(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 August 2002 (08.08.2002)

PCT

(10) International Publication Number
WO 02/061140 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US02/02892

(22) International Filing Date: 31 January 2002 (31.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/265,695 31 January 2001 (31.01.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/265,695 (CON)
Filed on 31 January 2001 (31.01.2001)

(71) Applicant *(for all designated States except US)*: **AMBION, INC.** [US/US]; Suite 200, 2130 Woodward Street, Austin, TX 78744 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BROWN, David** [US/US]; 1705 Drake Avenue, Austin, TX 78704 (US). **WINKLER, Matthew, M.** [US/US]; 960 Live Oak Circle, Austin, TX 78746 (US).

(74) Agent: **WILSON, Mark, B.**; Fulbright & Jaworski L.L.P., Suite 2400, 600 Congress Avenue, Austin, TX 78701 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/061140 A2

(54) Title: COMPETITIVE POPULATION NORMALIZATION FOR COMPARATIVE ANALYSIS OF NUCLEIC ACID SAMPLES

(57) Abstract: Methods are disclosed that convert two or more complex nucleic acid samples into a single collection of normalized target molecules that can be used to compare the abundance of each of the targets in the original samples. Multiple RNA or DNA samples are uniquely tagged and pooled to create a sample mixture. A defined set of target within the sample mixture is converted to approximately equal amounts of nucleic acid by one of several methods employing primer extension with a set of target specific primers. The concentration of target specific primers is equal and limiting for all targets, therefore an appropriate number of primer extension cycles converts all targets to similar final concentrations of product nucleic acid. The different tags appended to the sample nucleic acids from each sample population unique. These different tags are used to generate RNA or DNA , molecules for analysis that derive form each of the input samples. The disclosed methods are primarily intended to enhance gene array analysis, however, any method used to compare multiple nucleic acid targets form different samples will benefit form the invention.

DESCRIPTION

COMPETITIVE POPULATION NORMALIZATION FOR COMPARATIVE ANALYSIS OF NUCLEIC ACID SAMPLES

BACKGROUND OF THE INVENTION

This patent application claims priority to U. S. Provisional Patent Application No. 60/265,695.

The present application was filed concurrently with: PCT Application No. _____ on January 31, 2002, entitled "COMPARATIVE ANALYSIS OF NUCLEIC ACIDS USING POPULATION TAGGING," which claims priority to U.S. Provisional Patent Application No. 60/265,694, filed on January 31, 2001, PCT Application No. _____ filed January 31, 2002, entitled "METHODS FOR NUCLEIC ACID FINGERPRINT ANALYSIS," which claims priority to U. S. Provisional Patent Application No. 60/265,693 filed on January 31, 2001; and PCT Application No. _____, filed January 31, 2002 entitled "COMPETITIVE AMPLIFICATION OF FRACTIONATED TARGETS FROM MULTIPLE NUCLEIC ACID SAMPLES," which claims priority to U.S. Provisional Patent Application No. 60/265,692, filed on January 31, 2001. The disclosure of each of the above-identified applications is specifically incorporated herein by reference in its entirety without disclaimer.

1. Field of the Invention

The present invention relates generally to the fields of nucleic acid analysis. More particularly, it concerns methods for amplifying rare or limiting amounts of nucleic acid sequences for quantitative analysis.

2. Description of Related Art

Gene array analysis is a useful approach for comparing complex RNA and DNA populations. Gene arrays are solid supports upon which a collection of gene-specific probes has been spotted at defined locations. The probes localize complementary labeled targets from a nucleic acid sample population by hybridization. The amount of labeled target bound at each spot is quantified, providing a reasonable estimate of the abundance of the specific targets in the

population. Comparing the pattern of spots of one population to another provides insights into the molecular differences between the sample populations. Because the number of probes that can be spotted on a gene array is virtually unlimited, the arrays can be used to simultaneously screen thousands of targets in nucleic acid samples.

One of the most common uses for gene arrays is the comparison of the global expression patterns of different mRNA populations. A typical experiment involves isolating RNA from two or more tissue or cell samples. The RNAs are reverse transcribed using labeled nucleotides and target specific, oligodT, or random-sequence primers to create labeled cDNA populations. The cDNAs are denatured from the template RNA and hybridized to identical arrays. The hybridized signal on each array is detected and quantified. The signal emitting from each gene-specific spot is compared between the populations. Genes expressed at different levels in the samples generate different amounts of labeled cDNA and this results in spots on the array with different amounts of signal.

The direct conversion of RNA populations to labeled cDNAs is widely used because it is relatively simple and largely unaffected by enzymatic bias. However, direct labeling requires large quantities of RNA to create enough labeled product for moderately rare targets to be detected by array analysis. Most array protocols recommend that 2.5 μg of polyA or 50 μg of total RNA be used for reverse transcription (Duggan, 1999). For researchers unable to isolate this much RNA from their samples, global amplification procedures have been used.

The most often cited of these global amplification schemes is antisense RNA (aRNA) amplification (U.S. Patents 5,514,545 and 5,545,522, Phillips 1996). aRNA amplification involves reverse transcribing RNA samples with an oligo-dT primer that has a transcription promoter such as the T7 RNA polymerase consensus promoter sequence at its 5′ end. First strand reverse transcription creates single-stranded cDNA. Following first strand cDNA synthesis, the template RNA that is hybridized to the cDNA is partially degraded creating RNA primers. The RNA primers are then extended to create double-stranded DNAs possessing transcription promoters. The population is transcribed with an appropriate RNA polymerase to create an RNA population possessing sequence from the cDNA. Because transcription results in tens to thousands of RNAs being created from each DNA template, substantive amplification can

be achieved. The RNAs can be labeled during transcription and used directly for array analysis, or unlabeled aRNA can be reverse transcribed with labeled dNTPs to create a cDNA population for array hybridization. In either case, the detection and analysis of labeled targets via array analysis is the same as described above.

Whether RNA samples are labeled directly or used to amplify populations that are subsequently labeled, gene array methods rely on the direct detection of labeled nucleic acids. Modern detection methods have a threshold of detection of approximately 1,000,000 copies of an average labeled target (Lipshutz, 1999). In theory, all mRNA targets in a given population can be detected provided that the least abundant target is present at 1,000,000 copies or greater. The most straight forward way to accomplish this is to supply enough sample to the labeling or amplification reaction to generate at least 1,000,000 labeled molecules. However, increasing the amount of labeled product in an array hybridization reaction increases background. Background is the level of non-specific interactions between the labeled nucleic acids and the non-target probes on the array or the solid support upon which the gene probes are spotted. Background signal affects sensitivity and signal reliability, creating an upper limit to how much labeled sample can be applied to an array.

One study indicates that the threshold below which background exceeds signal occurs when a target RNA is present at less than 1 copy per 300,000 RNA molecules (Lockhart, 1996). This translates to limiting array analysis to mRNAs that are present at 3-30 copies per mammalian cell or more (Lockhart, 1996). While the instruments used for analyzing gene arrays will undoubtedly become more sensitive, it is unclear whether they will facilitate detection of less abundant targets than current instruments. Ultimately, the limitation of current methods owes to the presence of background signal affecting the specific signal from target molecules hybridized to complementary probes on the array. More sensitive instruments will be able to detect fewer than 1,000,000 labeled molecules, but they will also be affected by lower amounts of background. Therefore the detection limit for targets in a sample population might always be that the target be present in at least one copy per 300,000 molecules.

In theory, if all target molecules in a sample were present at the same concentration, then 300,000 targets could be assayed simultaneously (Lockhart 1996). However, the reality is that a

very small subset of messages (<20%) make up the majority of mRNA (75% or greater) present in a typical mammalian tissue sample (Zhang 1997). In fact, most of the mRNAs in a sample (86%) are expressed at less than five copies per cell (Zhang 1997). Existing labeling methods create labeled populations that reflect these ratios of abundant and rare RNA. Background caused by abundant targets or RNAs that are not even being assessed in the assay limits the detection of rare targets by creating background signal. The end result is that array experiments provide quantifiable signal from only 10-40% of the targets being studied when mammalian samples are being used. This greatly limits the effectiveness of arrays, as only a subset of targets can be studied.

If the relative abundance of the different targets could be normalized without impacting the ability to quantitatively compare the same targets between samples, then one could theoretically characterize any collection of targets within any set of samples regardless of their abundance.

SUMMARY OF THE INVENTION

In general, the invention relates to methods for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each potentially having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of at least a first target specific primer to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein any first product nucleic acids comprise the first differentiation domain and a section of the first nucleic acid target from the first sample and any second product nucleic acids comprise the second differentiation domain and a section of the first nucleic acid target from the second sample;

g) differentiating any first product nucleic acids from any second product nucleic acids; and

h) comparing the amount of first nucleic acid target in the first sample, if any, to the amount of first nucleic acid target of the second sample, if any.

It is important to recognize that the present invention is useful for determining the amount or abundance of a target nucleic acid in a sample, and that this encompasses the practice of the methods disclosed herein even when a target nucleic acid that is being assayed for is not present in a given sample. For example, it is possible that the target may be missing from a first sample, but present in a second sample in a given procedure. If this is the case, then it will not be possible to append a tag to the target in the first sample or, to amplify the target in the first sample. Therefore, the differentiation procedure will result in a determination that there was target present in the second sample, but not in the first. It is, therefore, not necessary that a target be present in any given sample for assays employing the methods disclosed herein to be within the scope of the invention.

In many applications, the nucleic acid target and/or the nucleic acid tag will be single-stranded nucleic acid. However this in not required in all embodiments of the invention, and those of skill will be able to follow the teachings of the specification to employ double-stranded nucleic acids in the invention. The nucleic acid target can be an RNA, DNA or a combination thereof. It is not required that the nucleic acid target be of natural origin, and the target can contain synthetic nucleotides. In specific aspects, the nucleic acid target is an RNA, for example, prokaryotic or eukaryotic RNA, total RNA, polyA RNA, an *in vitro* RNA transcript or a combination thereof. In other facets, the nucleic acid target may comprise DNA, such as, for example, cDNA , genomic DNA or a combination thereof. In certain aspects, at least one of the samples comprises nucleic acid isolated from a biological sample from, for example, a cell, tissue, organ or organism. In other aspects, at least one of the samples may comprise nucleic acid from an environmental sample. Of course, there is no need for all of the samples compared in a particular assay to be of the same source or type of source. A single sample may contain nucleic acid from a single source, or it may be the result of combining nucleic acids from multiple sources.

In some specific embodiments, the processing of the sample mixture further comprises amplification to produce amplified nucleic acid from the first sample and amplified nucleic acid from the second sample. For example, the amplification process may comprise a plurality of primer extension reactions.

In preferred embodiments, the first tag and the second tag both comprise amplification domains and a primer specific to the amplification domain is added to the extension reaction. In a specific embodiment, a non-limiting amount of a primer specific to the amplification domains and a limited amount of a primer specific to the first target are added to the sample mixture and the processing of the sample mixture comprises multiple cycles of primer extension to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample. The amplification domain of the first nucleic acid tag and the second nucleic acid tag may comprise a

primer binding domain. Alternatively, the amplification domain of the first nucleic acid tag and the second nucleic acid tag may comprise a transcription domain, or any other suitable domain. In preferred embodiments, the amplification domains of the first and second nucleic acid tags are functionally equivalent. And in even more preferred embodiments, the amplification domains of the first and second nucleic acid tags are identical.

While, at its most basic level, there can be only one nucleic acid of interest in the samples, the advantages of the invention allow one to analyze a variety of nucleic acid targets in the samples at the same time. Therefore, in many instances, the first nucleic acid target will be only one of a plurality of nucleic acid targets to be analyzed in the samples.

Further, while, at the most basic level, the methods of the invention may be employed with only two samples, in many cases, the first and second sample are two samples of a plurality of samples. One of the advantages of the invention is the ability of it to be used to analyze many samples simultaneously. In preferred embodiments, the tags used for each sample will comprise a differentiation domain that is unique to that sample.

Of course, in cases where there are a plurality of samples, there will typically be a plurality of tags. Those of skill in the art will be able to employ the teachings of this specification to prepare appropriate tags. Typically, the number of unique tags required for a given procedure will be equal to the number of samples to be analyzed. Also, the first target specific primer may be one of a plurality of target specific primers, each target specific primer present in a limiting concentration. In many cases, the first target specific primers anneal to the first target nucleic acid.

In some cases, processing comprises production of at least one first product nucleic acid complementary to the first nucleic acid target of the first sample.

In presently preferred embodiments of the invention, the differentiation domains of the tags are appended between the nucleic acid target sequence and the amplification domain. In this manner, the differentiation domain is assured of being amplified during the amplification process, and is present in the amplified nucleic acid. Of course, those of skill in the art will

realize that there are other positions of the differentiation and amplification domains in tags, and will be able to utilize tags with the domains in a variety of functional positions. In some preferred embodiments, the differentiation domain of the first nucleic acid tag or the second nucleic acid tag comprises a primer binding domain, a transcription domain, an affinity domain, a restriction enzyme cleavage domain, or a combination thereof.

In some specific embodiments, the differentiation domain of the first nucleic acid tag and the differentiation domain of the second nucleic acid tag each comprise a distinct primer binding domain, and differentiating comprises binding at least a first differentiation primer to at least one segment of the primer binding domain. In this case, differentiation may further comprise at least one primer extension reaction, which may produce at least one differentiated nucleic acid. In some cases, the differentiated nucleic acid is labeled with a detectable moiety.

In other specific embodiments, the differentiation domain of the first nucleic acid tag and the differentiation domain of the second nucleic acid tag each comprise distinct transcription domains. For example, the differentiation domain may comprise a prokaryotic RNA polymerase promoter. In some cases, differentiating comprises at least one transcription reaction, wherein the transcription reaction produces at least one differentiated nucleic acid. In such cases, the differentiated nucleic acid may be labeled with a detectable moiety.

In other specific embodiments, the differentiation domain of the first nucleic acid tag or the second nucleic acid tag comprises at least one affinity domain, and differentiating may comprise binding at least a first ligand to at least a segment of the affinity domain. In this case, the first ligand may comprise a nucleic acid, protein, or other biomolecule. The first ligand may be labeled with a detectable moeity. Also, the first ligand may be bound to at least one solid support, for example, an array, a microtiter well, a chip, a glass slide, a bead or a combination thereof. In some embodiments, the bound nucleic acids are removed from the first ligand.

In other embodiments, the differentiation domain of the first nucleic acid tag or the second nucleic acid tag comprises at least one restriction enzyme cleavage domain, and differentiation further comprising cleaving the restriction enzyme cleavage site to promote the ligation of a label to a segment of the at least a first or at least a second nucleic acid tag

sequence. In some cases, cleaving the restriction enzyme cleavage site promotes the ligation of at least one additional domain to a segment of the at least a first or at least a second nucleic acid tag sequence. Additionally, the differentiating may comprise cleaving the restriction enzyme site to remove at least one label.

In some embodiments, the first nucleic acid tag or the second nucleic acid tag further comprises at least one additional domain. For example, the additional domain can be a labeling domain, a restriction enzyme domain, a secondary amplification domain, a secondary differentiation domain or a combination thereof.

Some specific embodiments of the invention relate to methods for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of first target specific primers to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction with the mixed sample to produce a limited concentration of first product nucleic

acids complementary to the first nucleic acid target of the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, wherein the first product nucleic acids comprise the first differentiation domain and a section of the first nucleic acid target from the first sample and the second product nucleic acids comprise the second differentiation domain and a section of the first nucleic acid target from the second sample;

g) differentiating the first product nucleic acids from the second product nucleic acids; and

h) comparing the amount of first nucleic acid target in the first sample, to the amount of first nucleic acid target of the second sample.

Other specific embodiments relate to methods for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each potentially having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, wherein the first differentiation domain comprises a first transcription domain;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein the second differentiation domain comprises a second transcription domain that is different than the first transcription domain;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of first target specific primers to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein any first product nucleic acids comprise the first transcription domain and a section of the first nucleic acid target from the first sample and any second product nucleic acids comprise the second transcription domain and a section of the first nucleic acid target from the second sample;

g) differentiating any first product nucleic acids complementary to the first nucleic acid target of the first sample from any second product nucleic acids complementary to the first nucleic acid target of the second sample by a process comprising at least a first transcription reaction from the first transcription domain to produce an amount of at least first differentiated nucleic acid, if the first nucleic acid target is present in the first sample, and a second transcription reaction from the second transcription domain to produce at least second differentiated nucleic acid, if the first nucleic acid target is present in the second sample; and

h) comparing the amount of the first nucleic acid target in the first sample, if any, with the amount of the first nucleic acid target in the second sample, if any, by comparing the first differentiated nucleic acid, if any, to the amount of the second differentiated nucleic acid, if any.

Additional embodiments relate to methods for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each potentially having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, wherein the first differentiation domain comprises a first primer binding domain;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein the second differentiation domain comprises a second primer binding domain that is different than the first primer binding domain;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of first target specific primers to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein any first product nucleic acids comprise the first primer binding domain and a section of the first nucleic acid target from the first sample and any second product nucleic acids comprise the second primer binding domain and a section of the first nucleic acid target from the second sample;

g) differentiating any first product nucleic acids complementary to the first nucleic acid target of the first sample from any second product nucleic acids complementary to

the first nucleic acid target of the second sample by a process comprising annealing at least a first differentiation primer to the first primer binding domain and performing at least a second primer extension reaction to produce an amount of at least first differentiated nucleic acid, if the first nucleic acid target is present in the first sample, and annealing at least a second differentiation primer to the second primer binding domain and performing at least a third primer extension reaction to produce an amount of at least second differentiated nucleic acid, if the first nucleic acid target is present in the second sample; and

h) comparing the amount of the first nucleic acid target in the first sample, if any, with the amount of the first nucleic acid target in the second sample, if any, by comparing the first differentiated nucleic acid, if any, to the amount of the second differentiated nucleic acid, if any.

As used herein in the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one. As used herein "another" may mean at least a second or more. As used herein, a "plurality" means "two or more."

As used herein, "plurality" means more than one. In certain specific aspects, a plurality may mean 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 250, 300, 400, 500, 750, 1,000, 2,000, 3,000, 4,000, 5,000, 7,500, 10,000, 15,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 125,000, 150,000, 200,000 or more, and any integer derivable therein, and any range derivable therein.

As used herein, "any integer derivable therein" means a integer between the numbers described in the specification, and "any range derivable therein" means any range selected from such numbers or integers.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1. Configuration of Tag Domains Relative to the Target sequences to which they are Appended.

FIG. 2. NASBA/3SR-Mediated Nucleic Acid Amplification.

FIG. 3A and 3B. An Example of Competitive Population Normalization by Single Primer Extension.

FIG. 4A and 4B. An Example of Competitive Population Normalization Incorporating Nucleic Acid Amplification.

FIG. 5. Competitive Population Normalization Employing NASBA/3SR.

FIG. 6. Differentiation Via Unique Primer Binding Sites.

FIG. 7. Differentiation Via Unique Promoters for *In vitro* Transcription.

FIG. 8. Differentiation Via Unique Digestion Sites.

FIG. 9. Differentiation Via Unique Affinity Tags.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

To overcome the deficiencies of the art, the present invention provides novel methodologies that provide for the comparative analysis of targets regardless of their relative abundance in the populations being assayed.

One methodology of the invention allows for the creation of approximately equal amounts of nucleic acid from multiple targets in an RNA or DNA sample. The process of converting different nucleic acid targets present at differing concentrations in a sample to roughly equal amounts of labeled nucleic acids is referred to herein as "normalizing" or "target normalization." Target normalization can be applied to a variety of samples including total or polyA RNA or genomic DNA from tissues, cells, or organisms as well as environmental or diagnostic specimens. The number of unique target molecules that comprise the population being normalized can range from two to hundreds of thousands, though it is preferred that the method is used for studying tens to thousands of targets.

In target normalization, nucleic acid targets within a complex RNA or DNA sample are copied using equal and limiting concentrations of target specific primers in a primer extension reaction. The net effect is that all targets in the sample are copied to approximately the same concentration, providing normalization of the targets. Target normalization can be accomplished using any one of a variety of primer dependent methods to create sequences identical or complementary to a target RNA or DNA.

It is preferred that all reagents other than the target-specific primers that are required for copying the target nucleic acids are present in non-limiting amounts. Therefore primer extension proceeds until all or virtually all of the target specific primers have been exhausted. Because the target-specific primers are present in approximately equal concentrations, roughly equal numbers of copies of each target are generated to provide a normalized collection of target-specific sequences. Target normalization obviates the problems associated with abundant targets

generating enough labeled material to cause excess background signal in detection methods, such as those using arrays. Surprisingly, this technique is capable of generating approximately equal amounts of normalized nucleic acid from targets that vary in a sample population by as much as five orders of magnitude.

In a method that can function hand-in-hand with normalization in some embodiments, the invention allows for two or more nucleic acid samples to be mixed, applied to one or more enzymatic reactions, and analyzed such that the targets derived from each sample can be distinguished. The advantage to mixing and co-normalizing targets from different samples is that the conditions used for target normalization are shared by the target nucleic acids, making the assay more quantitative and less prone to tube-to-tube variability. The invention uses a nucleic acid tag that generally comprises at least one functional differentiation domain appended to the target nucleic acids in the samples being analyzed. The tags appended to each sample population are unique, therefore nucleic acids from each sample will share a sequence that distinguishes them from nucleic acids from other samples. The unique differentiation domains are most conveniently primer binding sites with unique sequences or transcriptional promoters for different RNA polymerases. Once different tags have been appended to the nucleic acid samples being investigated, the populations can be mixed and targets within the sample mixture can be normalized. Targets derived from each sample can be distinguished using the unique differentiation domains of the different tags.

In embodiments wherein the amount of each of the targets in the sample population exceeds the amount of target-specific primers, a single round of primer extension can be used to convert equal amounts of each of the targets into templates. In embodiments wherein one or more of the targets fail to saturate the level of target-specific primers, then multiple cycles of primer extension with both tag and target specific primers can be used to amplify each of the targets up to a similar concentration.

In certain embodiments, amplification is employed to copy the target nucleic acids. In these embodiments, it is particularly preferred that tags comprise an amplification domain in addition to a differentiation domain. In preferred aspects, tags comprising amplification domains that are the same for each of the populations being analyzed are appended to nucleic acids from

each sample. Identical primer binding sites for each target ensures that the targets are amplified from all samples with equal efficiency. FIG. 1 shows that tag sequences can be appended either 5' or 3' of a nucleic acid target. If amplification is being employed, an amplification domain can be included in the tag sequence, wherein it is particularly preferred that the differentiation domain lies between the target sequence and the amplification domain. In preferred aspects, the differentiation domains are used to distinguish the amplified nucleic acids derived from each of the populations. In further aspects, differentiated nucleic acids can be assessed by gene array analysis.

In certain embodiments, target normalization refers to the primer extension reactions that are used to convert a mixed population of tagged nucleic acid(s) to normalized nucleic acid(s) that can be used to generate differentiated nucleic acids for multi-target analysis. The term "processing" or "process" is used to account for the fact that single or multiple cycles of primer extension can be used to create a collection of normalized nucleic acid targets.

The two novel methodologies of sample tagging and target normalization have been incorporated into a single multi-step process that allows a collection of targets in two or more input nucleic acid samples to be directly compared regardless of the relative abundance of the targets in the original nucleic acid samples. In general embodiments, the resulting method comprises a tagging step that allows different nucleic acids from samples to be independently monitored after they are mixed with other samples and a normalization step that allows all of the targets being analyzed to be converted to roughly equivalent amounts of labeled nucleic acid. Target normalization is accomplished without affecting the relative abundance of any given target in one sample to the same target in another sample. This allows one to quantitatively compare the targets from one population to their counterparts in other populations. The benefit of normalization is that abundant targets do not create significant background, making it possible to simultaneously analyze abundant and rare targets.

The normalization of target concentration and the maintenance of the relative abundance of targets across different samples is referred herein as "competitive population normalization." In certain embodiments, competitive population normalization may be accomplished using a method comprising three steps: First, nucleic acid tags are appended to the members of the

different samples comprising RNA or DNA populations being studied. The tags are unique for each sample. Second, the tagged populations are mixed and at least a single cycle of primer extension is employed using approximately equal amounts of the target-specific primers that are the limiting reagent in the primer extension reaction. Third, the differentiation domains of the tags that are unique to each of the originial nucleic acid samples are used to generate nucleic acids from the sample mixture that derive from each of the original samples.

Using competitive population normalization, different samples that are being compared are differentially modified and combined into a single sample mixture. The sample mixture is then applied to a primer extension reaction using a limiting concentration of primers specific to the target sequences being compared. The extension reaction continues until the target specific primers are largely exhausted. The unique sequences appended to the sample populations are then used to generate labeled populations specific to each of the input sample populations.

The invention is compatible with single or multiple primer extension cycles. Single primer extension can be used wherein all target molecules in the mixed population exceed the detection threshold of the assay. Multiple primer extension cycles may be used to amplify target sequences wherein the sample is too small to provide detectable levels of all of the targets being compared.

A benefit of using the competitive population normalization methods described herein is that all targets present in the sample mixture will be detected, regardless of their abundance in the original samples. This derives from the fact that all targets generate an amount of RNA or DNA that exceeds the detection limit of the device being used for analysis but not enough to create excessive background. Therefore, unlike current sample preparation methods that generate enough labeled material from abundant targets to mask the signal of rare targets, the present invention generates detectable levels of all targets without creating enough labeled nucleic acids to cause background problems. Abundant targets are not over-represented because many of these cDNAs are not converted to double-stranded templates due to the limited availability of target-specific primers used for second strand synthesis. Non-target RNA sequences are not represented in the labeled sample applied to the array because cDNAs from the non-target RNAs are not converted to templates during the second-strand synthesis reaction.

Eliminating these two members from the labeled populations used for analysis reduces background and maximizes sensitivity, making it possible to detect every target represented on the array. In contrast, existing array protocols provide quantifiable signal from only 10-40% of the probe spots on a mammalian array.

A benefit of the present invention can be seen in regard to amplifying limited RNA samples. For example, in addition to providing a way to detect all targets in the samples, the methods of the present invention are more quantitative than other array based amplification schemes because it incorporates competitive amplification. Co-amplifying populations overcomes problems of amplification bias common to other methods that are being employed for array analysis. The nucleic acids being amplified are in exactly the same environment, thus any conditions that bias amplification of a given target are shared by the sample populations. In addition, each target from one population is competing for amplification reagents with its counterpart from the second population. Because the targets from the various populations are competing for the same pool of primers, it is preferred that the relative abundance of the target genes arising from each pool should be maintained until the gene specific primers are exhausted.

The present invention uses non-standard concentrations of target specific primers to normalize targets within a sample. Other methods have utilized non-standard concentrations of primers for PCR amplification. For example, the methods described in U.S. Patent 6,015,664 or Henegariu *et al.,* 1997 use non-standard primer concentrations to reduce amplification efficiency for defined primers. In certain embodiments, the present invention uses gene specific primers at limited concentrations to facilitate amplification of many different targets. Amplification efficiency is largely irrelevant in the methods of the present invention, as an unlimited number of amplification cycles can be used to ensure that all targets are completely amplified. In contrast, the prior art references do not seem to discuss limiting primer concentrations to allow all targets to be amplified to detectable levels regardless of their relative abundance within the population. Rather non-standard concentrations of PCR primers are used to modulate amplification efficiency or primer use.

The use of tag sequences to co-amplify target sequences from two or more populations has been described to assess single targets in multiple complex nucleic acid populations using a

technique called Adaptor-Tagged Competitive-PCR (ATAC-PCR) (Kato 1997, European Patent Application No. 98302726). ATAC-PCR involves generating double-stranded cDNA from multiple RNA populations. The cDNAs are digested with a restriction enzyme and unique adapters are appended to the cDNAs via ligation at the restriction site. The adapters share a primer binding site but differ in size or sequence (*i.e.*, unique restriction or hybridization sites). The adapter-tagged cDNAs are mixed and PCR™ amplified with a gene-specific primer and a primer specific to the shared adapter sequence present at the proximal ends of the cDNA populations. If the adapters used for tagging were different sizes, then the PCR™ products are directly assessed by gel electrophoresis. If the adapters from the populations differ by a restriction site, then the PCR™ amplified population must be aliquoted into different restriction digestion reactions to create products that can be distinguished by gel electrophoresis. Because the products arising from the populations are different sizes, they can be readily fractionated and quantified.

The present invention is different from ATAC-PCR in a variety of ways. In certain embodiments, the methods of the present invention are designed to amplify multiple targets in a single reaction to facilitate parallel quantitative analysis of nucleic acids. In contrast, ATAC-PCR is described and applied to assess single targets in multiple populations using PCR™ amplification with single tag and target-specific primers (Japanese patent, Kato 1997, Matoba 2000).

In other embodiments, the methods of the present invention employ limiting amounts of target specific primers to generate equal concentrations of many different target-specific templates. In contrast, ATAC-PCR uses standard amplification conditions to maximize the amplification efficiency of single targets.

In some embodiments, the present invention provides differential labeling of nucleic acids derived from different sample populations. ATAC-PCR does not use a differentiation reaction to create labeled nucleic acids derived from different samples for analysis. Instead, it directly assesses the amplification products by gel electrophoresis. Differential labeling by a differentiation domains of the tagged samples, allowing samples to be assessed by array analysis.

Additionally, in certain embodiments the present invention is compatible with a variety of nucleic acid amplification methods as well as being capable of being used independent of any nucleic acid amplification. In contrast, ATAC-PCR depends on PCR™ to generate DNA fragments for analysis.

A. NUCLEIC ACIDS: TAGS AND SAMPLES

Embodiments of the present invention involve nucleic acids in many forms. Nucleic acid samples are collections of RNA and/or DNA derived or extracted from chemical or enzymatic reactions, biological samples, or environmental samples. Nucleic acid tags are nucleic acids of a defined sequence that are appended to nucleic acids in a sample to facilitate its analysis. There are many potential types of tags for use in the invention, which are described elsewhere in this specification.

1. General Description of Nucleic Acids

The general term "nucleic acid" is well known in the art. A "nucleic acid" as used herein will generally refer to a molecule (*i.e.*, a strand) of DNA, RNA or a derivative or analog thereof, comprising a nucleobase. A nucleobase includes, for example, a naturally occurring purine or pyrimidine base found in DNA (*e.g.*, an adenine "A," a guanine "G," a thymine "T" or a cytosine "C") or RNA (*e.g.*, an A, a G, an uracil "U" or a C). The term "nucleic acid" encompasses the terms "oligonucleotide" and "polynucleotide," each as a subgenus of the term "nucleic acid." The term "oligonucleotide" refers to a molecule of between 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, and 100 nucleobases in length, and any range derivable therein. The term "polynucleotide" refers to at least one molecule of greater than about 100 nucleobases in length.

a. Nucleobases

As used herein a "nucleobase" refers to a heterocyclic base, such as for example a naturally occurring nucleobase (*i.e.*, an A, T, G, C or U) found in at least one naturally occurring nucleic acid (*i.e.*, DNA and RNA), and naturally or non-naturally occurring derivative(s) and

analogs of such a nucleobase. A nucleobase generally can form one or more hydrogen bonds ("anneal" or "hybridize") with at least one naturally occurring nucleobase in a manner that may substitute for naturally occurring nucleobase pairing (*e.g.*, the hydrogen bonding between A and T, G and C, and A and U).

"Purine" and/or "pyrimidine" nucleobase(s) encompass naturally occurring purine and/or pyrimidine nucleobases and also derivative(s) and analog(s) thereof, including but not limited to, a purine or pyrimidine substituted by one or more of an alkyl, caboxyalkyl, amino, hydroxyl, halogen (*i.e.*, fluoro, chloro, bromo, or iodo), thiol or alkylthiol moeity. Preferred alkyl (*e.g.*, alkyl, caboxyalkyl, etc.) moeities comprise of from about 1, about 2, about 3, about 4, about 5, to about 6 carbon atoms. Other non-limiting examples of a purine or pyrimidine include a deazapurine, a 2,6-diaminopurine, a 5-fluorouracil, a xanthine, a hypoxanthine, a 8-bromoguanine, a 8-chloroguanine, a bromothymine, a 8-aminoguanine, a 8-hydroxyguanine, a 8-methylguanine, a 8-thioguanine, an azaguanine, a 2-aminopurine, a 5-ethylcytosine, a 5-methylcyosine, a 5-bromouracil, a 5-ethyluracil, a 5-iodouracil, a 5-chlorouracil, a 5-propyluracil, a thiouracil, a 2-methyladenine, a methylthioadenine, a N,N-diemethyladenine, an azaadenines, a 8-bromoadenine, a 8-hydroxyadenine, a 6-hydroxyaminopurine, a 6-thiopurine, a 4-(6-aminohexyl/cytosine), and the like. A table of non-limiting purine and pyrimidine derivatives and analogs is also provided herein below.

Table 1-Purine and Pyrmidine Derivatives or Analogs

Abbr.	Modified base description	Abbr.	Modified base description
Ac4c	4-acetylcytidine	Mam5s2u	5-methoxyaminomethyl-2-thiouridine
Chm5u	5-(carboxyhydroxylmethyl) uridine	Man q	Beta,D-mannosylqueosine
Cm	2'-O-methylcytidine	Mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
Cmnm5s2u	5-carboxymethylamino-methyl-2-thioridine	Mcm5u	5-methoxycarbonylmethyluridine
Cmnm5u	5-carboxymethylaminomethyluridine	Mo5u	5-methoxyuridine
D	Dihydrouridine	Ms2i6a	2-methylthio-N6-isopentenyladenosine
Fm	2'-O-methylpseudouridine	Ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
Gal q	Beta,D-galactosylqueosine	Mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methyl-carbamoyl)threonine
Gm	2'-O-methylguanosine	Mv	Uridine-5-oxyacetic acid methylester
I	Inosine	O5u	Uridine-5-oxyacetic acid (v)
I6a	N6-isopentenyladenosine	Osyw	Wybutoxosine
M1a	1-methyladenosine	P	Pseudouridine
M1f	1-methylpseudouridine	Q	Queosine
M1g	1-methylguanosine	s2c	2-thiocytidine

Table 1-Purine and Pyrmidine Derivatives or Analogs

Abbr.	Modified base description	Abbr.	Modified base description
M1I	1-methylinosine	s2t	5-methyl-2-thiouridine
M22g	2,2-dimethylguanosine	s2u	2-thiouridine
M2a	2-methyladenosine	s4u	4-thiouridine
M2g	2-methylguanosine	T	5-methyluridine
M3c	3-methylcytidine	t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)carbamoyl)threonine
M5c	5-methylcytidine	Tm	2'-O-methyl-5-methyluridine
M6a	N6-methyladenosine	Um	2'-O-methyluridine
M7g	7-methylguanosine	Yw	Wybutosine
Mam5u	5-methylaminomethyluridine	X	3-(3-amino-3-carboxypropyl)uridine, (acp3)u

A nucleobase may be comprised in a nucleoside or nucleotide, using any chemical or natural synthesis method described herein or known to one of ordinary skill in the art.

b. Nucleosides

As used herein, a "nucleoside" refers to an individual chemical unit comprising a nucleobase covalently attached to a nucleobase linker moiety. A non-limiting example of a "nucleobase linker moiety" is a sugar comprising 5-carbon atoms (*i.e.*, a "5-carbon sugar"), including but not limited to a deoxyribose, a ribose, an arabinose, or a derivative or an analog of a 5-carbon sugar. Non-limiting examples of a derivative or an analog of a 5-carbon sugar include a 2'-fluoro-2'-deoxyribose or a carbocyclic sugar where a carbon is substituted for an oxygen atom in the sugar ring.

Different types of covalent attachment(s) of a nucleobase to a nucleobase linker moiety are known in the art. By way of non-limiting example, a nucleoside comprising a purine (*i.e.*, A or G) or a 7-deazapurine nucleobase typically covalently attaches the 9 position of a purine or a 7-deazapurine to the 1'-position of a 5-carbon sugar. In another non-limiting example, a nucleoside comprising a pyrimidine nucleobase (*i.e.*, C, T or U) typically covalently attaches a 1 position of a pyrimidine to a 1'-position of a 5-carbon sugar (Kornberg and Baker, 1992).

c. Nucleotides

As used herein, a "nucleotide" refers to a nucleoside further comprising a "backbone moiety". A backbone moiety generally covalently attaches a nucleotide to another molecule comprising a nucleotide, or to another nucleotide to form a nucleic acid. The "backbone moiety" in naturally occurring nucleotides typically comprises a phosphorus moiety, which is covalently attached to a 5-carbon sugar. The attachment of the backbone moiety typically occurs at either the 3'- or 5'-position of the 5-carbon sugar. However, other types of attachments are known in the art, particularly when a nucleotide comprises derivatives or analogs of a naturally occurring 5-carbon sugar or phosphorus moiety.

d. Nucleic Acid Analogs

A tag or other nucleic acid used in the invention may comprise, or be composed entirely of, a derivative or analog of a nucleobase, a nucleobase linker moiety and/or backbone moiety that may be present in a naturally occurring nucleic acid. As used herein a "derivative" refers to a chemically modified or altered form of a naturally occurring molecule, while the terms "mimic" or "analog" refer to a molecule that may or may not structurally resemble a naturally occurring molecule or moiety, but possesses similar functions. As used herein, a "moiety" generally refers to a smaller chemical or molecular component of a larger chemical or molecular structure. Nucleobase, nucleoside and nucleotide analogs or derivatives are well known in the art, and have been described (see for example, Scheit, 1980, incorporated herein by reference).

Additional non-limiting examples of nucleosides, nucleotides or nucleic acids comprising 5-carbon sugar and/or backbone moiety derivatives or analogs, include those in U.S. Patent No. 5,681,947 which describes oligonucleotides comprising purine derivatives that form triple helixes with and/or prevent expression of dsDNA; U.S. Patents 5,652,099 and 5,763,167 which describe nucleic acids incorporating fluorescent analogs of nucleosides found in DNA or RNA, particularly for use as fluorescent nucleic acids probes; U.S. Patent 5,614,617 which describes oligonucleotide analogs with substitutions on pyrimidine rings that possess enhanced nuclease stability; U.S. Patents 5,670,663, 5,872,232 and 5,859,221 which describe oligonucleotide analogs with modified 5-carbon sugars (*i.e.*, modified 2'-deoxyfuranosyl moieties) used in nucleic acid detection; U.S. Patent 5,446,137 which describes oligonucleotides comprising at least one 5-carbon sugar moiety substituted at the 4' position with a substituent other than hydrogen that can be used in hybridization assays; U.S. Patent 5,886,165 which describes oligonucleotides with both deoxyribonucleotides with 3'-5' internucleotide linkages and ribonucleotides with 2'-5' internucleotide linkages; U.S. Patent 5,714,606 which describes a modified internucleotide linkage wherein a 3'-position oxygen of the internucleotide linkage is replaced by a carbon to enhance the nuclease resistance of nucleic acids; U.S. Patent 5,672,697 which describes oligonucleotides containing one or more 5' methylene phosphonate internucleotide linkages that enhance nuclease resistance; U.S. Patents 5,466,786 and 5,792,847 which describe the linkage of a substituent moeity which may comprise a drug or label to the 2' carbon of an oligonucleotide to provide enhanced nuclease stability; U.S. Patent 5,223,618 which describes oligonucleotide analogs with a 2 or 3 carbon backbone linkage attaching the 4'

position and 3' position of adjacent 5-carbon sugar moiety to enhanced resistance to nucleases and hybridization to target RNA; U.S. Patent 5,470,967 which describes oligonucleotides comprising at least one sulfamate or sulfamide internucleotide linkage that are useful as nucleic acid hybridization probe; U.S. Patents 5,378,825, 5,777,092, 5,623,070, 5,610,289 and 5,602,240 which describe oligonucleotides with three or four atom linker moeity replacing phosphodiester backbone moeity used for improved nuclease resistance; U.S. Patent 5,214,136 which describes olignucleotides conjugated to anthraquinone at the 5' terminus that possess enhanced hybridization to DNA or RNA; enhanced stability to nucleases; U.S. Patent 5,700,922 which describes PNA-DNA-PNA chimeras wherein the DNA comprises 2'-deoxy-erythro-pentofuranosyl nucleotides for enhanced nuclease resistance and binding affinity; and U.S. Patent 5,708,154 which describes RNA linked to a DNA to form a DNA-RNA hybrid.

e. Polyether and Peptide Nucleic Acids

In certain embodiments, it is contemplated that a tag or other nucleic acid comprising a derivative or analog of a nucleoside or nucleotide may be used in the methods and compositions of the invention. A non-limiting example is a "polyether nucleic acid", described in U.S. Patent Serial No. 5,908,845, incorporated herein by reference. In a polyether nucleic acid, one or more nucleobases are linked to chiral carbon atoms in a polyether backbone.

Another non-limiting example is a "peptide nucleic acid", also known as a "PNA", "peptide-based nucleic acid analog" or "PENAM", described in U.S. Patent Serial Nos. 5,786,461, 5891,625, 5,773,571, 5,766,855, 5,736,336, 5,719,262, 5,714,331, 5,539,082, and WO 92/20702, each of which is incorporated herein by reference. Peptide nucleic acids generally have enhanced sequence specificity, binding properties, and resistance to enzymatic degradation in comparison to molecules such as DNA and RNA (Egholm *et al.*, 1993; PCT/EP/01219). A peptide nucleic acid generally comprises one or more nucleotides or nucleosides that comprise a nucleobase moiety, a nucleobase linker moeity that is not a 5-carbon sugar, and/or a backbone moiety that is not a phosphate backbone moiety. Examples of nucleobase linker moieties described for PNAs include aza nitrogen atoms, amido and/or ureido tethers (see for example, U.S. Patent No. 5,539,082). Examples of backbone moieties described for PNAs include an aminoethylglycine, polyamide, polyethyl, polythioamide, polysulfinamide or polysulfonamide backbone moiety.

In certain embodiments, a nucleic acid analogue such as a peptide nucleic acid may be used to inhibit nucleic acid amplification, such as in PCR, to reduce false positives and discriminate between single base mutants, as described in U.S. Patent Serial No. 5891,625. Other modifications and uses of nucleic acid analogs are known in the art, and are encompassed by the invention. In a non-limiting example, U.S. Patent 5,786,461 describes PNAs with amino acid side chains attached to the PNA backbone to enhance solubility of the molecule. Another example is described in U.S. Patent Nos. 5,766,855, 5,719,262, 5,714,331 and 5,736,336, which describe PNAs comprising naturally and non-naturally occurring nucleobases and alkylamine side chains that provide improvements in sequence specificity, solubility and/or binding affinity relative to a naturally occurring nucleic acid.

f. Preparation of Nucleic Acids

A tag or other nucleic acid used in the invention may be made by any technique known to one of ordinary skill in the art, such as for example, chemical synthesis, enzymatic production or biological production. Non-limiting examples of a synthetic nucleic acid (*e.g.*, a synthetic oligonucleotide), include a nucleic acid made by *in vitro* chemical synthesis using phosphotriester, phosphite or phosphoramidite chemistry and solid phase techniques such as described in EP 266,032, incorporated herein by reference, or via deoxynucleoside H-phosphonate intermediates as described by Froehler *et al.*, 1986 and U.S. Patent Serial No. 5,705,629, each incorporated herein by reference. In the methods of the present invention, one or more oligonucleotides are used. Various different mechanisms of oligonucleotide synthesis have been disclosed in for example, U.S. Patents. 4,659,774, 4,816,571, 5,141,813, 5,264,566, 4,959,463, 5,428,148, 5,554,744, 5,574,146, 5,602,244, each of which is incorporated herein by reference.

A non-limiting example of an enzymatically produced nucleic acid includes one produced by enzymes in amplification reactions such as PCR (see for example, U.S. Patent 4,683,202 and U.S. Patent 4,682,195, each incorporated herein by reference), or the synthesis of an oligonucleotide described in U.S. Patent No. 5,645,897, incorporated herein by reference. A non-limiting example of a biologically produced nucleic acid includes a recombinant nucleic

acid produced (*i.e.*, replicated) in a living cell, such as a recombinant DNA vector replicated in bacteria (see for example, Sambrook *et al.* 1989, incorporated herein by reference).

g. Nucleic Acid Purification

A tag or other nucleic acid used in the invention may be purified on polyacrylamide gels, cesium chloride centrifugation gradients, or by any other means known to one of ordinary skill in the art (see for example, Sambrook *et al.* 1989, incorporated herein by reference).

In particular embodiments, tags or other nucleic acid used in the invention may be isolated from at least one organelle, cell, tissue or organism. In certain embodiments, "isolated nucleic acid" refers to a nucleic acid that has been isolated free of, or is otherwise free of, the bulk of cellular components such as for example, macromolecules such as lipids or proteins, small biological molecules, and the like.

h. Nucleic Acid Complements

The present invention also encompasses a nucleic acid that is complementary to a specific nucleic acid sequence. A nucleic acid "complement(s)" or is "complementary" to another nucleic acid when it is capable of base-pairing with another nucleic acid according to the standard Watson-Crick, Hoogsteen or reverse Hoogsteen binding complementarily rules. As used herein "another nucleic acid" may refer to a separate molecule or a spatial separated sequence of the same molecule.

i. Hybridization

As used herein, "hybridization", "hybridizes" or "capable of hybridizing" is understood to mean the forming of a double or triple stranded molecule or a molecule with partial double or triple stranded nature. The term "anneal" as used herein is synonymous with "hybridize." The term "hybridization", "hybridize(s)" or "capable of hybridizing" encompasses the terms "stringent condition(s)" or "high stringency" and the terms "low stringency" or "low stringency condition(s)."

As used herein "stringent condition(s)" or "high stringency" are those conditions that allow hybridization between or within one or more nucleic acid strand(s) containing

complementary sequence(s), but precludes hybridization of random sequences. Stringent conditions tolerate little, if any, mismatch between a nucleic acid and a target strand. Such conditions are well known to those of ordinary skill in the art, and are preferred for applications requiring high selectivity. Non-limiting applications include isolating a nucleic acid, such as a gene or a nucleic acid segment thereof, or detecting at least one specific mRNA transcript or a nucleic acid segment thereof, and the like.

Stringent conditions may comprise low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleobase content of the target sequence(s), the charge composition of the nucleic acid(s), and to the presence or concentration of formamide, tetramethylammonium chloride or other solvent(s) in a hybridization mixture.

It is also understood that these ranges, compositions and conditions for hybridization are mentioned by way of non-limiting examples only, and that the desired stringency for a particular hybridization reaction is often determined empirically by comparison to one or more positive or negative controls. Depending on the application envisioned it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of a nucleic acid towards a target sequence. In a non-limiting example, identification or isolation of a related target nucleic acid that does not hybridize to a nucleic acid under stringent conditions may be achieved by hybridization at low temperature and/or high ionic strength. Such conditions are termed "low stringency" or "low stringency conditions", and non-limiting examples of low stringency include hybridization performed at about 0.15 M to about 0.9 M NaCl at a temperature range of about 20°C to about 50°C. Of course, it is within the skill of one in the art to further modify the low or high stringency conditions to suit a particular application.

B. NUCLEIC ACID SAMPLES (POPULATIONS)

The invention can be applied to the comparative analysis of any nucleic acid population. The nucleic acids can be RNA, DNA, or both. The nucleic acids can be part of a collection of other molecules, including proteins, carbohydrates or small molecules. While the population can

comprise even a single sequence, the method is best suited for nucleic acid samples that include hundreds or thousands of unique sequences.

The terms "target", "target nucleic acid" and "target sequence" refer to one or more nucleic acids (*e.g.*, DNA, RNA) of a specific sequence that are being characterized. Often, target nucleic acids comprise a sub-population of nucleic acids relative to all the nucleic acid sequences originally present in a nucleic acid sample.

1. Sources of Nucleic Acid Samples

Nucleic acid samples can be obtained from biological material, such as cells, tissues, organs or organisms. The invention is particularly relevant to total and polyA RNA preparations from tissues or cells. Similarly, the invention could be applied to cDNAs derived from cells or tissues. In other embodiments, multiple genomic DNA samples could be assessed using the methods of the present invention.

a. Cells and Tissues

A cell, or a tissue comprising cells, may be a source of nucleic acids for the present invention.. In certain embodiments, cells or tissue may be part of or separated from an organism. In certain embodiments, a cell or tissue may comprise, but is not limited to, adipocytes, alveolar, ameloblasts, axon, basal cells, blood (*e.g.*, lymphocytes), blood vessel, bone, bone marrow, brain, breast, cartilage, cervix, colon, cornea, embryonic, endometrium, endothelial, epithelial, esophagus, facia, fibroblast, follicular, ganglion cells, glial cells, goblet cells, kidney, liver, lung, lymph node, muscle, neuron, ovaries, pancreas, peripheral blood, prostate, skin, skin, small intestine, spleen, stem cells, stomach, testes, anthers, ascites, cobs, ears, flowers, husks, kernels, leaves, meristematic cells, pollen, root tips, roots, silk, stalks, and all cancers thereof.

b. Organisms

In certain embodiments, an organism may be a source of nucleic acids for the present invention. In certain embodiments, the organism may be, but is not limited to, a eubacteria, an archaea, a eukaryote or a virus (for example, webpage http://phylogeny.arizona.edu/tree/phylogeny.html).

i. Eubacteria

In certain embodiments, the organism is a eubacteria. In particular embodiments, the eubacteria may be, but is not limited to, an aquifecales; a thermotogales; a thermodesulfobacterium; a member of the thermus-deinococcus group; a chloroflecales; a cyanobacteria; a firmicutes; a member of the leptospirillum group; a synergistes; a member of the chlorobium-flavobacteria group; a member of the chlamydia-verrucomicrobia group, including but not limited to a verrucomicrobia or a chlamydia; a planctomycetales; a flexistipes; a member of the fibrobacter group; a spirochetes; a proteobacteria, including but not limited to an alpha proteobacteria, a beta proteobacteria, a delta & epsilon proteobacteria or a gamma proteobacteria. In certain aspects, an organelle derived from eubacteria are contemplated, including a mitochondria or a chloroplast.

ii. Archaea

In certain embodiments, the organism is an archaea (a.k.a. archaebacteria; *e.g.,* a methanogens, a halophiles, a sulfolobus). In particular embodiments, the archaea may be, but is not limited to, a korarchaeota; a crenarchaeota, including but not limited to, a thermofilum, a pyrobaculum, a thermoproteus, a sulfolobus, a metallosphaera, an acidianus, a thermodiscus, a igneococcus, a thermosphaera, a desulfurococcus, a staphylothermus, a pyrolobus, a hyperthermus or a pyrodictium; or an euryarchaeota, including but not limited to a halobacteriales, methanomicrobiales, a methanobacteriales, a methanococcales, a methanopyrales, an archeoglobales, a thermoplasmales or a thermococcales.

iii. Eukaryotes

In certain embodiments, the organism is a eukaryote (*e.g.,* a protist, a plant, a fungi, an animal). In particular embodiments, the eukaryote may be, but is not limited to, a microsporidia, a diplomonad, an oxymonad, a retortamonad, a parabasalid, a pelobiont, an entamoebae or a mitochondrial eukaryote (*e.g.,* an animal, a plant, a fungi, a stramenopiles).

iv. Viruses

In certain embodiments the organism may be a virus. In particular aspects, the virus may be, but is not limited to, a DNA virus, including but not limited to a ssDNA virus or a dsDNA virus; a DNA RNA rev transcribing virus; a RNA virus, including but not limited to a dsRNA

virus, including but not limited to a -ve stranded ssRNA or a +ve stranded ssRNA; or an unassigned virus.

c. Synthetic Samples

Nucleic acid samples comprising populations designed by the hand of man may also be generated and used as a standard against which another sample or subpopulation of target sequences could be compared. The synthetic population can be used to accurately quantify one or more targets from one or more sample(s) if the concentrations of the synthetic nucleic acids are known. For example, a synthetic sample may comprise a collection of nucleic acids (*e.g.*, RNA, cDNA or genomic DNA) from many different tissues, cells (*e.g.*, cell cultures), or other samples that could provide an average population against which a sample, or subpopulation of target sequences, could be compared. In another non-limiting example, the synthetic sample could comprise a collection of *in vitro* transcripts at known or unknown concentrations sharing a specific tag sequence so that they could be co-amplified with nucleic acids from another sample (*e.g.*, RNA) to quantify a collection of targets. In another example, the synthetic sample could comprise a set of DNAs at known or unknown concentrations sharing a specific tag sequence that could be used to quantify a sample comprising a target DNA population.

d. Sample Mixtures

A sample mixture is a collection of two or more nucleic acid samples (*e.g.*, RNA, cDNA or DNA). It is particularly preferred that the different nucleic acid samples that comprise the sample mixture are distinguishable. This is typically achieved by differentially tagging the targets of each input sample prior to mixing.

C. TAGS

The invention involves appending a tag to one or more target sequences, up to all nucleic acid sequences, present in a nucleic acid sample. A tag is a common sequence shared by various nucleic acid sequences of a sample that allows nucleic acids of one sample to be distinguished from nucleic acids from another sample. The term tag is also used to describe the RNA, DNA, or other nucleic acid molecule appended or otherwise used to attach a tag sequence to the nucleic acids comprising a sample. In preferred embodiments, a tag is an RNA, DNA, or other molecule that can be used as a template by a polymerase to generate a complementary strand.

A tag comprises at least one functional domain, referred to as a "differentiation domain", that can be used to distinguish the nucleic acid target(s) derived from each sample in a sample mixture. Where amplification is being applied, the tag sequence generally comprises a second functional element referred to as an "amplification domain." Thus, in preferred embodiments, a tag comprises at least two functional domains, an amplification domain compatible with amplification and a differentiation domain that can be used to distinguish nucleic acids that derive from the samples being assessed. Of course, a tag may comprise one or more additional sequences. Generally, additional sequences will possess functional properties, such as, for example, a property that facilitates analysis of amplified and/or differentiated nucleic acids.

An example of a general makeup of a tag comprising an amplification and differentiation domain is provided in FIG. 1. Whether the tag includes one or multiple functional domains, a differentiation domain should be adjacent to the portion of the nucleic acid target sequences being assessed. It is particularly preferred that the differentiation domain is between an amplification domain and the sequence of each target nucleic acid of the population to be amplified. In other words, it is particularly preferred that a differentiation domain is internal to the amplification domain.

Although they are drawn as distinct domains (FIG. 1), the differentiation and amplification domain sequences can overlap, though it is particularly preferred that they are functionally distinct. In other words, it is particularly preferred that the amplification domain does not co-function as the primary differentiation domain in the tag. For example, it is particularly preferred that the presence of the amplification domain sequence alone is not sufficient to allow differentiation of the amplified nucleic acids for analysis.

1. Amplification Domains

In most embodiments, it is particularly preferred that a tag comprise at least one amplification domain. As used herein, an amplification domain will primarily be a sequence that can support the amplification of a nucleic acid that includes such sequence. Use of nucleic acid sequences in amplification reactions are well known in the art, and non-limiting examples are described herein.

In certain embodiments, an amplification domain will comprise a sequence that can support primer binding and extension. Standard rules for primer design apply (Sambrook, 1994). In specific aspects, an amplification domain will preferably comprise a primer binding site for PCR™ amplification. PCR™ does not require any specialized structure or sequence to sustain amplification; the PCR™ amplification primer typically contains only binding sequences. Parameters for primer design for PCR™ are well known in the art (see, *e.g.*, Beasley *et al.*, 1999).

Primer binding sites for other types of amplification methods might also be used in amplification domains. Often such primer binding regions share similar characteristics with PCR™ primer binding sites, however the primers used for other amplification methods typically possess sequences 5' to the binding domain. For instance, primers for 3SR and NASBA contain an RNA polymerase promoter sequence 5' to the priming site to support subsequent transcription (FIG. 2). Because 3SR and NASBA are performed at relatively low temperature (37°C to 42°C), the amplification domains can have much lower melting temperatures than those used for PCR™.

In embodiments wherein a sample mixture is assessed, the amplification domains of the tags used in the samples that were mixed will preferably be identical to facilitate equal co-amplification of the target sequences from the different input samples that comprise the mixed sample being assessed.

2. Differentiation Domains

It is particularly preferred that a tag comprise at least one differentiation domain. A differentiation domain comprises a sequence that can be used to identify the sample from which a particular amplified nucleic acid derives. For example, a differentiation domain may comprise an affinity sequence for removing one or more nucleic acids from a sample mixture that derive from a particular sample, a primer binding site for sample specific labeled DNA synthesis, or a different transcription promoter for sample specific labeled RNA synthesis.

a. **Affinity Domains**

In certain embodiments, a differentiation domain may provide an affinity site for hybridization or binding (an "affinity domain") to a ligand comprising, but not limited to, a nucleic acid, protein or other molecule.

For example, an affinity domain may be used to extract from an amplified population one or more target nucleic acids derived from a given sample. In certain aspects, the affinity site may comprise a sequence that can hybridize to a nucleic acid ligand (*e.g.*, an oligonucleotide or polynucleotide). In certain facets, the sample specific ligands may be bound to different solid supports that can be incubated with the sample population and then removed to provide target nucleic acids derived from single sample populations.

b. **Primer Binding Domains**

A differentiation domain may comprise a primer binding site (a "primer binding domain"). A primer binding site may provide an annealing site for various types of primers that can be extended by a polymerase to generate a labeled DNA. Binding sites for primers are well known in the art (Sambrook 1989)

c. **Transcription Domains**

In certain embodiments, a differentiation domain may comprise a promoter sequence (a "transcription domain") that binds an RNA polymerase to initiate transcription. In certain embodiments, a DNA in a sample mixture can be transcribed to generate labeled RNA derived from a single sample. For example, nucleic acids possessing promoter sequences can be transcribed in a reaction (*e.g.*, an *in vitro* reaction) with one or more labeled nucleotides (radio- or non-isotopic-labeled NTPs) and an appropriate RNA polymerase to convert double-stranded DNA into differentiated RNAs that can be used for comparative analysis.

3. **Additional Functional Domains**

A tag may comprise one or more additional functional or structural sequences in addition to the primary amplification and differentiation domains, as described herein or as would be known to one of ordinary skill in the art. In certain embodiments, these domains may be partly or fully comprised within other domains, such as, for example an amplification domain or a

differentiation domain. In other embodiments, these additional domains may be comprised in sequences that do not comprise the amplification domain or differentiation domain.

These additional domain(s) may be used to support additional molecular biological reactions, including but not limited to, a nested amplification reaction, a secondary differentiation reaction, a labeling reaction, a hybridization reaction or a combination thereof. The addition of one or more additional domains will be particularly preferred in certain embodiments for manipulating one or more samples, including the nucleic acid samples that comprise a sample mixture.

Additional sequences described herein are by no means intended as an exhaustive list of all of the potential functional domains that can be included to facilitate production, amplification, differentiation, comparison or analysis of a nucleic acid target, or sample. The list is merely intended to provide examples of some requirements and benefits of additional functional domains that can be incorporated into the nucleic acid tag.

a. Labeling Domains

A tag may comprise a sequence that is used in a labeling reaction (a "labeling domain") to convert nucleic acids synthesized via target-specific primer extension into a labeled product population for subsequent differentiation and analysis. A variety of sequences can be used to support the production of labeled products, and non-limiting examples are described herein. In specific embodiments, a labeling domain may be used for labeled DNA or labeled RNA product synthesis. It is particularly preferred that the labeling domain be situated upstream of the differentiation domain so that the labeled nucleic acids include the differentiation domain sequence.

b. Secondary Amplification Domains

One or more additional amplification domains may be used for nested amplification(U.S. Patent 5,340728). In general embodiments, nested amplification comprises sequential amplification reactions wherein a first amplification with a first set of one or more primers generates one or more primary amplified nucleic acids, and an at least a second amplification of the one or more primary amplified nucleic acids with another set of primers comprising a primer

that binds a sequence partly or fully internal to a primer of the first set, so that a nucleic acid segment of the one or more primary amplified nucleic acids is then amplified to produce one or more secondary amplified nucleic acids. In certain embodiments, nested amplification might be required for those targets that are present in only a few copies in a sample or where small amounts of a sample (*e.g.*, a few mammalian cells) are available. In certain embodiments, a secondary amplification domain may be located between the primary amplification domain and the primary differentiation domain.

c. Secondary Differentiation Domains

One or more additional differentiation domains may be used in conjunction with the primary differentiation domain to further distinguish a population. For example, if transcription is being used to differentiate the samples that comprise a sample mixture, and only a few different polymerases are available for *in vitro* transcription, then only a few input samples can be assayed at a time. Incorporating a secondary differentiation domain into the tag would allow additional input samples' populations to be mixed and assayed by the methods of the present invention.

For example, incorporating a primer binding site between the amplification domain and the primary differentiation domain would allow additional sample populations to be mixed. In one aspect, several samples could use tags with the same transcription promoter that comprises their primary differentiation domain so long as their secondary differentiation domains were unique. The primary amplification would use a single tag-specific primer for all samples. The amplified population could then be split and further amplified with primers specific to the secondary differentiation domains. Each of the resulting samples could then be used to generate differentiated populations for analysis, such as, for example, using the different transcription promoters.

D. METHODS FOR APPENDING TAGS TO NUCLEIC ACID POPULATIONS

A nucleic acid tag of the present invention may be added to or appended to a nucleic acid population. As would be appreciated by one of ordinary skill in the art, different methods of tag attachment or incorporation may be used depending on whether the nucleic acid population comprises DNA or RNA. Non-limiting examples of such methods that may be used are

described herein, though other methods can be used as would be known by one of ordinary skill in the art.

In addition to the techniques described herein, other methods of nucleic acid manipulation and/or additional compositions may be applied to nucleic acid targets, populations and/or samples. Such additional methods include for example, but are not limited to, separation and/or detection. Such additional methods and compositions are described in detail in U.S. Patent Application No. 60/265,694, entitled "COMPARATIVE ANALYSIS OF NUCLEIC ACIDS USING POPULATION TAGGING," filed on January 31, 2001; U.S. Patent Application No. 60/265,692, entitled "COMPETITIVE AMPLIFICATION OF FRACTIONATED TARGETS FROM MULTIPLE NUCLEIC ACID SAMPLES," filed on January 31, 2001; and U.S. Patent Application No. 60/265,693, entitled "METHODS FOR NUCLEIC ACID FINGERPRINT ANALYSIS," filed on January 31, 2001; and each of whose disclosure is specifically incorporated herein by reference in their entirety without disclaimer.

1. Tagging RNA

The methods of the present invention are applicable to tagging both eukaryotic RNA and/or prokaryotic RNA. In other aspects, the present invention may be applied to tag polyA selected or total RNA populations. As will be apparent to one of ordinary skill in the art in light of the disclosures herein, a tag may be appended to RNA populations in a variety of ways. Non-limiting examples of methods of tagging RNA are described below.

Once an RNA molecule is tagged, it can undergo further molecular biology reactions, including but not limited to reverse transcription and amplification. In preferred embodiments, reverse transcription and amplification can be accomplished using sequences present in the ligated tag. For example, a tagged RNA sample may be mixed with other tagged RNA samples, converted to cDNA and the cDNA amplified or differentiated with at least one primer specific to the tag. If targets in the cDNA are to be amplified, then primers specific to both the tag and target(s) need to be used.

a. Ligation

In certain embodiments, a tag can be appended to the 3' ends of RNAs by a ligase (*e.g.*, an enzymatic protein, nucleic acid or chemical that induces ligation). For ligation, an excess of RNA or DNA polynucleotide tag possessing a 5' phosphate can be added to a RNA population. Incubation of the mixture with a ligating agent (*e.g.*, RNA ligase) generates RNAs with the tag ligated to the 3' end of the RNAs.

In general embodiments, more efficient ligation may be achieved by adding bridging oligonucleotides to the ligation reaction. Hybridization of a bridge to both an RNA to be tagged and a tag will align the 3' and 5' ends of the two molecules, enhancing ligation efficiency. In a non-limiting example, a bridging oligonucleotide may comprise a sequence at its 3' end that is complementary to the 3' end of the RNAs in the sample and a sequence at its 5' end that is complementary to the 5' end of the tag.

b. Cap Dependent Ligation

In one embodiment, a cap dependent ligation may be used to selectively append tags to the 5' ends of eukaryotic mRNAs. In general aspects, an RNA may be tagged by the combined enzymatic activities of a phosphatase (*e.g.*, calf intestinal phosphatase), a pyrophosphatase (*e.g.*, tobacco acid pyrophosphatase) that leaves a 5' phosphate at the 5' terminus of a capped message, and nucleic acid ligase (*e.g.*, RNA ligase).

In a non-limiting example, a total RNA population is treated with calf intestinal phophatase (CIP) to dephosphorylate the RNA population. CIP is specific to RNAs with free terminal phosphates, therefore the 5' phosphates of rRNAs, tRNAs, and partially degraded mRNAs are removed leaving these RNAs with 5' hydroxyls. After the CIP is inactivated, the RNA preparation is treated with a phosphatase such as tobacco acid pyrophoshatase (TAP) to convert the 5' cap structures of mRNAs to 5' monophosphates. An excess of a DNA or RNA polynucleotide tag is added to the RNA population as well as a ligase that functions on RNA substrates. The tag should ligate exclusively to TAP modified RNAs possessing 5' monophosphates as all of the non-capped RNAs possess 5' hydroxyls following CIP treatment. The resulting tagged mRNA population can be used in subsequent reactions for comparative analysis.

c. **Enzymatic Polymerization**

In an additional embodiment, a tag is incorporated into an RNA population by enzymatic polymerization. A tag comprising a 3' nucleotide that cannot be extended by polymerization (see for example, U.S. Patent 6,057,134), can be hybridized to the 3' ends of an RNA population. An RNA or DNA polymerase with the ability to extend primer template junctions can be added to the mixture and allowed to extend the 3' ends of the RNAs in the population, incorporating a sequence complementary to the hybridized oligonucleotide at the 3' ends of the RNA in the sample. If the nucleic acid that serves as a template comprises a tag sequence, then the polymerization reaction effectively tags the RNA within the sample.

d. **CAPswitch™**

A method for tagging mRNAs by Cap-induced primer extension is described in U.S. Patent 5,962,271. The technology, referred to as CAPswitch™, uses a unique CAPswitch oligonucleotide in the first strand cDNA synthesis reaction. When reverse transcriptase stops at the 5' end of an mRNA template in the course of first strand cDNA synthesis, it switches to a CAPswitch oligonucleotide and continues DNA synthesis to the end of a CAPswitch oligonucleotide. The resulting cDNA has at its 3' end a sequence that is complementary to the CAPswitch oligonucleotide sequence. The CAPswitch technology may be used to tag one or more RNA populations by using one or more CAPswitch oligonucleotides comprising differentiation or differentiation and amplification domains.

2. **Tagging RNA Populations by Reverse Transcription**

In a preferred embodiment, tag sequences may be appended to sample nucleic acids by reverse transcription. For example, tagged cDNA populations can be conveniently generated by priming reverse transcription with oligonucleotide tags comprising sequence at their 3′ end that is complementary to a specific sequence of one or more targets in a RNA sample. Of course, it is preferred that such tags comprise a differentiation domain 5′ to the domain that hybridizes to the RNA sample. In certain other preferred aspects, the tags comprise a differentiation and an amplification domain at their 5′ ends.

In a particularly preferred example, eukaryotic mRNAs that possess a polyA tail can be tagged using a primer that has a polyT or polyU at its 3′ end and appropriate tag sequences 5′ to the polyT or polyU. The polyA specific tag primer hybridizes to and is extended from the polyA tail of the mRNAs. The resulting cDNAs possess the tag sequences at or near their 5′ ends. The tag domains can be used in subsequent reactions to facilitate comparative analysis.

3. Tagging Prokaryotic RNA Samples

The methods described above may be less preferred for tagging a non-polyA RNA, such as a prokaryotic RNA. However, analysis of prokaryotic RNA samples is desirable in certain aspects.

In certain embodiments, it may be desirable to remove or separate small RNAs (*e.g.*, tRNAs) from prokaryotic mRNA or any other RNA population lacking a polyA tail. Methods of removing small RNA are known to those of skill in the art, and include such methods as a lithium chloride precipitation. Lithium chloride precipitation is specific to RNAs greater than 100-300 nucleotides, thus tRNAs and other small RNAs will be removed from the RNA population (Sambrook 1989).

In embodiments wherein non-rRNAs are the target RNAs, one or more rRNA specific oligonucleotides and a polyA polymerase can be used to add polyA tails to a target RNA. For example, a sample comprising prokaryotic total RNA can be precipitated with lithium chloride. After removal of the solution comprising the tRNAs, a resulting RNA population can be resuspended and hybridized to one or more oligonucleotides complementary to the 3' ends of the major prokaryotic rRNAs. The 5' ends of the oligonucleotide(s) will preferably extend beyond the 3' ends of the rRNAs, creating a slight 5' overhang. The RNA population can then be treated with a polyA polymerase and ATP. RNAs with non-hybridized 3' ends can be extended by the action of the polymerase, creating a 3' polyA tail on the mRNA portion of the sample. The resulting polyA modified RNA can then be reverse transcribed using a tag comprising a oligo-dT target specific domain sequence.

4. Tagging DNA

DNA (*e.g.*, genomic DNA and cDNA) can be tagged by various methods, including primer extension or ligation. In certain aspects, the DNA may be single stranded or double stranded.

a. Single Stranded DNA

In one embodiment, a single-stranded DNA (*e.g.*, cDNA) may be diluted in a buffer appropriate for hybridization and polymerization, and hybridized to tags possessing sequences at their 3′ end complementary to DNA in the sample. Addition of a DNA polymerase such as, for example, the klenow fragment of DNA polymerase I or Taq DNA polymerase, will extend a tag to create a tagged population of DNA segments.

In aspects where the DNA is double stranded (*e.g.*, genomic DNA), it may be denatured prior to tagging by any of a variety of methods known in the art, including, for example, heating to 95°C in a solution of 0.2 M NaOH. In certain aspects, the denatured DNA may be removed or purified from the denaturing reagents by methods well known to those of skill in the art, such as, for example, ethanol precipitation. The denatured DNA may then be tagged using primer extension as described above or any other techniques that would be known to one of ordinary skill in the art.

b. Double Stranded DNA

In certain embodiments, double stranded DNA may be tagged by ligation. For example, a double-stranded DNA can be digested with a restriction enzyme, and one or more tags comprising a compatible restriction fragment cut site may be ligated to the digested DNA.

A disadvantage of appending double-stranded tags to double-stranded nucleic acids (*e.g.*, DNA) is that when amplification is being used to process nucleic acid samples, the primers specific to the amplification domain of the tag can bind and be extended from target and non-target molecules alike. Using restriction digestion and double-stranded tag ligation may create far greater background than the other methods and is therefore a less preferred method for tagging populations. This is in contrast to other preferred tagging methods described herein, whereby single-stranded tags are appended to single-stranded nucleic acids from the sample. In

the embodiments whereby differentially tagged samples comprise single-stranded nucleic acids, the amplification domain of the tag sequence only becomes a primer binding site when the target specific primer is extended during the amplification phase.

E. COMPETITIVE POPULATION NORMALIZATION

Maximizing the yield of target nucleic acids in a sample mixture is important. However, the amount of each target nucleic acid will be limited to ensure that all targets can be produced to detectable levels. These two goals are achieved by providing sufficient nucleotides, polymerase, and tag primers to the reaction to maximize overall yield, while limiting the concentration of each gene specific primer in the reaction to a level that exceeds the detection limit of the assay but whose sum is less than the second most limiting reagent in the normalization (*e.g.*, primer extension and/or amplification) reaction. For example, if ten targets are being analyzed where the second most limiting reagent is the tag-specific amplification primer at 1 μM, the total concentration of the ten target-specific primers would total 1 μM or less. Although not a requirement, the concentrations of the ten target-specific primers would preferably be equal, meaning that each primer would be present at a concentration of not more than 0.1 μM.

A particularly preferred element of the invention is the use of target-specific primer concentrations that normalize the level of targets in a nucleic acid population comprising multiple samples. Although it is not absolutely required, the concentrations of the target-specific primers will often be approximately equal. The factors affecting the concentration of target-specific primer being used for a particular extension and/or amplification reaction include: the detection limit of the instrument being used for analysis; the number of samples being normalized; the method used for differentiation; and the number of targets being assessed.

1. Analytical Instruments

The detection limit of the analytical instrument provides a lower limit for the amount of detectable nucleic acid needed for each target from each sample population. It is contemplated that this will dictate the lower limit of target-specific primers. For example, if the instrument being used has a detection limit of 1,000,000 labeled nucleic acids, then a minimum of 1,000,000 primers for each target will be required for target normalization. In practical terms, 10,000,000

primers will likely be required for each target to ensure that targets from all of the input samples will be converted to at least one million labeled nucleic acid molecules.

2. Sample Numbers

The number of samples being normalized may also influence the concentration of the target specific primers, as more samples will require more primers to ensure that targets from all of the populations exceed the detection threshold.

3. Differentiation Methods

The method used for differentiation may affect the number of target-specific primers needed for converting target nucleic acids into differentiable products. If transcription is being used for differentiation, then multiple RNAs will typically be generated from each template. Fewer templates will be required to provide enough nucleic acids from each target for analysis, allowing fewer primers to be used for sample processing. Primer extension and direct detection of the normalized samples will require more gene specific primers to ensure detection.

The theoretical limit of how many targets can be characterized from a sample mixture is equal to the difference between the overall yield of labeled target nucleic acids achievable in a series of normalization and differentiation reactions and the detection threshold for the analysis system. For example, assuming an overall PCR™ yield of 1 μM, based on the concentration of tag-specific primers in a 50 μl reaction and equal concentrations of target-specific primers, 3 x 10^7 different targets could be amplified up to the 1,000,000 molecule detection limit of existing array analysis systems. The breakdown for this calculation is that 50 μl of a 1 μM solution of all target-specific primers can produce up to 3×10^{13} total DNA molecules. 3×10^{13} molecules divided by the 10^6 detection limit for each target provides the 3×10^7 different targets that could theoretically be amplified up to 1,000,000 copies. This many targets exceed by two orders of magnitude the 10^5 genes that are believed to comprise the human genome. This means it may be possible to compare the relative abundance of every RNA in two or more human tissue or cell samples using the methods of the present invention, regardless of how rare they are in the samples.

It is contemplated that certain factors may affect the maximum efficiency of the present invention. First, labeled nucleic acids for each target from each population must exceed the detection limit to provide meaningful quantitative data. Therefore, the minimal concentration for each target-specific primer should preferably exceed the detection limit by ten to one-hundred-fold. If amplification is being employed, those targets that have exhausted their supply of target specific primers will still be bound and extended by the tag-specific primers during subsequent primer extension cycles. This constant removal of tag-specific primers will reduce the effective concentration later in the amplification reaction and thus limit the amplification of the most rare targets. It is contemplated that providing a ten to one-hundred-fold excess of tag-specific primer above the theoretical yield of the target-specific nucleic acids will likely ensure that the supply of tag primers will not be exhausted. Thus, it is contemplated that the methods of the present invention may allow 10^3 to 10^4 targets to be amplified to 1,000,000 copies in a 50 µl competitive population normalization reaction.

It is also contemplated that the number of targets that can be evaluated from a single normalizing reaction will improve if the differential labeling reactions can generate multiple labeled nucleic acids from each nucleic acid resulting from the normalization reaction. For example, if labeling is achieved by transcription, DNA templates will typically support the synthesis of 10 to 1,000 labeled RNAs. Thus, if transcription is used for the differentiated RNA, the number of target-specific nucleic acids resulting from the normalization reaction necessary for detection can be reduced by at least an order of magnitude, increasing the number of targets that can be detected by an order of magnitude.

4. Competitive Population Normalization by Single Primer Extension

In one embodiment, the methods of the present invention are applied to samples that are large enough to provide each of the target nucleic acids at a level that facilitates detection (1,000,000 labeled molecules, for example). This is shown in FIG. 3.

In such an embodiment, a first nucleic acid tag comprising a differentiation domain having a first transcription domain (i.e., a T7 promoter) is appended (via reverse transcription) to a plurality of RNA targets of at least a first nucleic acid population. A second nucleic acid tag comprising a differentiation domain having a second transcription domain (i.e., a SP6 promoter)

is appended (via reverse transcription) to a plurality of RNA targets of at least a second input nucleic acid population. Although many forms of differentiation domains may be used in various embodiments, the differentiation domain of the second nucleic acid tag is preferably different than the differentiation domain of the first nucleic acid tag. The two nucleic acid populations are mixed to produce a sample mixture.

The sample mixture is then admixed with a collection of target specific primers present at approximately equal concentrations that exceed the threshold of detection for the device that is to be used for analysis. The tagged target(s) in the sample mixture are annealed to and co-extended with one or more target specific primers, wherein the concentration of the target specific primer(s) are limited (*e.g.*, 1,000,000 copies each)and the other reagents are not limited, producing normalized nucleic acids (*e.g.*, 1,000,000 double stranded cDNAs per target). These normalized nucleic acids comprise at least a differentiation domain of the first nucleic acid tag or second nucleic acid tag and a nucleic acid segment of the targets.

The double-strand DNA resulting from target primer extension is split into two differentiation reactions, one with T7 RNA polymerase and one with SP6 RNA polymerase. The reaction with T7 RNA polymerase creates approximately 20 RNAs per double-stranded cDNA derived from the first RNA sample that possess a T7 RNA polymerase promoter. The SP6 reaction generates similar yields of labeled RNA from target cDNAs derived the second RNA sample. Different labeling agents (*e.g.*, Cy5 UTP and Cy3 UTP) may be used to produce labeled nucleic acids (i.e., labeled RNA).

The differentiated nucleic acids can be bound to a plurality of target specific ligands on one or more solid support(s) (*e.g.*, arrays) for comparison of the relative abundance of the targets in the original populations.

5. Competitive Population Normalization Using Amplification

Most nucleic acid samples being analyzed are available in limited quantities. For these samples, the vast majority of interesting target nucleic acids fall below the detection threshold of current instruments, limiting the usefulness of gene arrays. Competitive population normalization can be used to amplify all target sequences above the threshold of detection for the

instrument being used without being dominated by the abundant targets or non-targets in the sample. A sample protocol is shown in FIG. 4.

RNA samples are reverse transcribed in separate reactions with different tags possessing the following general structure:

Amplification Domain **Differentiation Domain**

Primer Binding Domain Transcription Domain

5' ——————————————————— $(dT)_{20}VN$ 3'.

The first sample uses a reverse transcription primer that comprises a differentiation domain comprising a first transcription domain (i.e., a T7 promoter), an amplification domain (i.e., a primer binding domain), and anchored oligodT at the 3' end of the primer. Reverse transcription of the second RNA sample is primed by an anchored oligodT primer possessing a differentiation domain comprising a second transcription domain (i.e., a SP6 promoter), and an amplification domain (FIG. 4). The primer binding sites for both tags are typically identical.

The tagged cDNAs are mixed to create a sample mixture. The tagged target(s) in the sample mixture are annealed to and co-amplified with one or more target specific primers and a primer to the amplification domain, wherein the concentration of the target specific primer(s) are limited (i.e., 1,000,000 copies each) and the other reagents are not limited, producing normalized nucleic acids (i.e., 1,000,000 double stranded cDNAs per target) comprising at least a differentiation domain of the first nucleic acid tag or second nucleic acid tag and a nucleic acid segment of the targets (FIG. 4).

The normalized nucleic acids are separated into different transcription reactions, one with T7 RNA polymerase; one with SP6 RNA polyerase. The reaction with T7 RNA polymerase creates approximately 20 RNAs per double-stranded cDNA derived from targets that possess a T7 RNA polymerase promoter (targets from the first RNA sample). The SP6 reaction generates similar yields of labeled RNA from target cDNAs derived from the second RNA sample. Different labeling agents (*e.g.*, Cy5 UTP and Cy3 UTP) may be incorporated during transcription. The resulting labeled RNAs can be bound to a plurality of target specific ligands on one or more solid support(s) (*e.g.*, arrays) for comparison of the relative abundance of the targets in the original populations.

F. AMPLIFICATION

In certain preferred embodiments of the invention, nucleic acid amplification is employed to convert sample populations into a collection of differentiable target sequences. In these

embodiments, the tag sequences must include an amplification domain. A variety of methods have been described for nucleic acid amplification, and are known to those of skill in the art. It is particularly preferred that an amplification step used in a method of the present invention require target specific primers. In preferred embodiments, target specific primers are used to normalize concentrations of the various targets.

Most amplification techniques have proven to be incompatible with generating amplified nucleic acids for array analysis, because the amplification efficiencies of different target sequences vary. Efficiently amplified targets tend to dominate the amplification reactions making poorly amplified targets difficult or even impossible to detect. In addition, the amplification efficiencies of each target is affected by the makeup of the sample population, therefore a target might be amplified more efficiently in one population than in another. These two properties create amplification biases that make quantitative comparative analysis impossible. Comparative population normalization overcomes these problems by limiting the amplification of all targets to some threshold and co-amplifying all samples in a single reaction so that amplification biases are shared by all of the samples being compared.

For amplification reactions, it is preferred to remove any unincorporated tags prior to amplification to keep the unincorporated tags from competing with amplification primers for templates during amplification. An oligonucleotide can be removed from the sample using, for example, size exclusion chromatography (Sambrook 1989). In one embodiment, supports with a pore size large enough to allow the tags to enter while excluding the larger nucleic acids provides an easy way to generate primer-free nucleic acids. In some embodiments, the free tags can be removed from a nucleic acid population by differential precipitation. For example, LiCl and ethanol are both known to preferentially precipitate larger DNA, therefore, as would be known to one of ordinary skill in the art, appropriate conditions may be developed to separate DNA from the oligonucleotide tags prior to amplification.

Although slight variations abound, the general principle of nucleic acid amplification is the same. In embodiments for amplification of DNA, a sample comprising a DNA population is contacted with one or more amplification primers that are able to hybridize to targets comprising the DNA population with amplification reagents in appropriate amplification conditions.

Once hybridized to complementary domains in the sample nucleic acids, the primer(s) are extended by a nucleic acid polymerase (*e.g.*, a DNA polymerase) to create copies of the target nucleic acids in the sample mixture. Repeated cycles of primer hybridization and polymerase extension leads to the exponential accumulation of normalized nucleic acids. Any such nucleic acid amplification techniques can be used in the context of the invention.

As described in this application, the preferred amplification method for the invention is the polymerase chain reaction (referred to as PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, 1988. Each of these references is incorporated herein by reference in their entirety. Other non-limiting methods for amplification of target nucleic acid sequences that may be used in the practice of the present invention are disclosed in U.S. Patent Nos. 5,843,650, 5,846,709, 5,846,783, 5,849,546, 5,849,497, 5,849,547, 5,858,652, 5,866,366, 5,916,776, 5,922,574, 5,928,905, 5,928,906, 5,932,451, 5,935,825, 5,939,291 and 5,942,391, GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety.

In another embodiment, a reverse transcriptase PCR™ amplification procedure may be performed to amplify mRNA populations. Methods of reverse transcribing RNA into cDNA are well known (see Sambrook, 1989). Alternative methods for reverse transcription utilize thermostable DNA polymerases. These methods are described in WO 90/07641. Additionally, representative methods of RT-PCR are described in U.S. Patent No. 5,882,864. RT-PCR could be applied to tagged RNA samples described herein to amplify and normalize target RNAs in a sample mixture.

Other non-limiting nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh *et al.*, 1989; Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference in their entirety). European Application No. 329 822 disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention.

Nucleic Acid Sequence Based Amplification (NASBA) (Guatelli, 1990; Compton, 1991) makes use of three enzymes, avian myeloblastosis virus reverse transcriptase (AMV-RT), *E. coli* RNase H, and T7 RNA polymerase to induce repeated cycles of reverse transcription and RNA transcription.

As shown in FIG. 2, the NASBA reaction begins with the priming of first strand cDNA synthesis with a gene specific oligonucleotide (primer 1) comprising a T7 RNA polymerase promoter. The first primer anneals to the target and is extended by reverse transcriptase (RT). The primer includes a domain complementary to the target mRNA (polyT) and a promoter for T7 RNA polymerase. The RNA target is partially degraded by RNAse H, and a second primer (primer 2) anneals to the DNA strand and is extended by RT, yielding a double-stranded DNA template with an active T7 promoter region. T7 RNA polymerase binds to this promoter and transcribes the template strand, yielding 100 to 1000 RNA copies. Primer 2 anneals to each of these RNA copies and is extended by RT, and the RNA strand is degraded by RNAse H. The promoter-primer then binds to the DNA strand, and extension with RT produces the dsDNA template with an active T7 promoter. More RNA copies are made, and this cycle continues in an autocatalytic fashion.

Competitive population normalization using NASBA/3SR can employ uniquely tagged samples and co-amplification with a limiting concentration of target-specific primer and an excess of tag-specific primer (FIG. 5).

As shown in FIG. 5, multiple nucleic acid populations are differentially tagged with sequences that include a T7 RNA polymerase promoter (T7) 5' to a NASBA primer binding site (PBS in Primers 1 and 2) that is 5' to a differentiation domain (DD#1 in Primer 1; DD#2 in Primer 2). A differentiation domain can be adjacent to the oligodT site and a primer binding site for regenerating the transcriptional promoter in cDNAs generated from RNA transcripts can be adjacent to the transcription promoter at the 5' end of the originating RT tag (FIG. 5).

The tagged populations of cDNA are mixed and converted to double-stranded DNA using target-specific primers (Primer 4) that are present at equal and limiting concentrations using NASBA. For example, a limiting amount of target specific primers (*e.g.*, 100 pM) and an

excess of an RT primer with a promoter at its 5′ end and sequence complementary to the 5′ ends of NASBA/3SR transcripts can be used to amplify and normalize target sequences. The target-specific primers bind targets from the mixed cDNA population, providing primers for reverse transcription. The initial double-stranded cDNA population will support transcription from the promoter at the 5′ end of the tag. The target specific primers hybridize to the resulting transcripts and are extended by the action of AMV RT. The tag specific primers hybridize to the 3′ ends of the cDNAs and can be extended by the reverse transcriptase. The double-stranded cDNAs will support additional transcription. Amplification of the double-stranded cDNA continues until the target-specific primers are exhausted, at which point, all target templates are present at roughly the same concentration.

The amplification reaction could then be aliquoted into multiple differentiation reactions specific to the double-stranded cDNAs of amplification, and appropriate differentiation reactions used to create labeled nucleic acids for multi-target analysis. The labeled nucleic acids resulting from the differentiation reactions can be used to compare multiple targets for the nucleic acid samples.

G. DIFFERENTIATION REACTIONS

A differentiation reaction may comprise any of a number of methods that distinguish from which sample a nucleic acid derives. In preferred embodiments, differentiation comprises any reaction that generates differentiated nucleic acids for analysis that derive exclusively from one sample within a sample mixture. A differentiated nucleic acid generally is created through transcription, primer extension, enzymatic modification, or binding through the differentiation domain of the tag sequence to allow ready detection and/or quantification.

In certain embodiments, it is preferred that the differentiation domains only become functional after they are copied by the extension of target specific primers. This ensures that the limited concentration of the target specific primers creates a normalized population of templates. Differentiation domains may comprise, for example, an initiation domain for DNA or RNA synthesis, affinity sites for extracting specific labeled members of the amplified population, or a digestion domains for removing a label used for detection or nucleic acid purification.

1. Differential Labeled Nucleic Acid Synthesis

In a preferred embodiment, a differentiation domain is used to generate one or more labeled products for analysis. As used herein, a "labeled product" or a "labeled nucleic acid" is a nucleic acid that comprises a detectable labeling molecule or moiety (a "labeling agent"). Labeling agents include non-isotopic reagents, isotopic reagents or combinations thereof. Non-isotopic compounds used for labeling are typically an affinity ligand such as, for example, a biotin, a digoxigenin, a DNP or fluoresent dye such as Cy3 or Cy5 that are attached covalently to a primer, one or more dNTPs being incorporated or both. Alternatively, one or more radiolabeled dNTPs (*e.g.*, α-^{32}P, α -^{33}P, or α -^{35}S dNTPs) may be incorporated into the primer, dNTPs, or both. Of course, other labeling agents that would be known to those of skill in the art in light of the disclosures herein may be used.

i. Primer Extension

Labeled nucleic acids may be created by using a differentiation domain that comprises a binding site for a differentiation primer. In this case, multiple nucleic acid populations can be differentiated using tag sequences with different primer binding sites for each nucleic acid sample.

An example of this is shown in FIG. 6, where a first nucleic acid tag comprising a differentiation domain that comprises a first primer binding domain (PBS#1) is appended to a first nucleic acid target of a first nucleic acid population, and a second nucleic acid tag comprising a differentiation domain with a second primer binding domain (PBS#2) is appended to the first nucleic acid target of a second nucleic acid population In this example, the differentiation domain of the second nucleic acid tag is different than the differentiation domain of the first nucleic acid tag. In FIG. 6, as elsewhere in the figures, the thick lines represent the tag sequences, and the thin lines represent the sequences of the RNA and/or DNA populations in which one or more nucleic acid targets are comprised.

The differentially tagged samples are mixed to create a sample mixture (FIG. 6). Tagged target(s) in the sample mixture anneal to one or more target specific primers, wherein the concentration of the target specific primer(s) are limited, producing at least a first normalized nucleic acid comprising at least a differentiation domain of the first nucleic acid tag and a

nucleic acid segment of the target(s), and at least a second normalized nucleic acid comprising at least a differentiation domain of the second nucleic acid tag and a nucleic acid segment of the target(s). Multiple cycles of primer extension can be used to increase the number of copies of the target. The normalized target nucleic acids are separated into differentiation reactions employing binding primers specific to the different primer binding domains of the different differentiation domains.

Labeled DNA is produced by extending the annealed primer using a polymerase and incorporating a labeled nucleotide or by using a labeled primer.

In certain embodiments, a differentiation primer may be hybridized to the binding site and extended by a DNA polymerase (*e.g.*, klenow fragment of DNA polymerase I or Taq DNA polymerase) to produce the labeled nucleic acid. Labeled primers are well known in the art, and include, for example, a primer possessing a 5' terminal radiolabeled phosphate.

A benefit of using labeled differentiation primers is that a single reaction can be used to differentially label one or more targets from one or more (*e.g.*, all) of the sample populations. For example, if the differentiation primer used to label targets from one sample has Cy3 and the differentiation primer for the target of the second sample has Cy5, then the two primers can be added to the nucleic acids resulting from target normalization. The differentiation primers will hybridize to the appropriate templates and be extended by the action of a DNA polymerase. Targets derived from one sample will be labeled exclusively with Cy3 while targets from the second sample will be labeled with Cy5. The labeled differentiated nucleic acids can then be added to an array and allowed to hybridize. Target detection can be performed in a way that the signals from Cy3 and Cy5 can be distinguished, providing a measure of the relative abundance of each of the targets in the input samples' populations (Chee 1996).

Where labeled nucleotides are being used to create labeled, differentiated DNA for comparative analysis, different primer extension reactions are used to create labeled products specific to each sample. In each reaction, a differentiation primer specific to a single sample is extended by a polymerase incorporating labeled nucleotides. The labeled nucleic acids generated in the differentiated reaction can then be used for comparative analysis.

In embodiments wherein labeled nucleotides are being incorporated, such as, for example, using the primer extension method, it is preferred to keep an amplification primer from being extended during the differentiation reaction. Because the primers used prior to differentiation can hybridize equally well to all of the sample populations, the labeled products resulting from the extension of any non-differentiation primers would be as likely to derive from an unintended sample as an intended sample. The labeled product would therefore not be specific to a single sample making the labeled products incompatible with comparative analysis. Thus, in particularly preferred aspects, the non-differentiation primers are removed prior to initiating a differentiation reaction. A primer can be removed using techniques that would be known to those of skill in the art, such as for example, size exclusion chromatography or precipitation of nucleic acids using conditions that keep primers in solution (Sambrook, 1996). For example, a nucleic acid population can be added to a size exclusion column and centrifuged. The population collects in the filtrate, free of the column-bound amplification primers.

ii. Differential Labeling by *In vitro* Transcription

In embodiments wherein tags include a differentiation domain that is a transcription promoter, a transcription reaction with one or more labeled nucleotides (*e.g.*, isotopic- or non-isotopic-labeled NTPs) and an appropriate RNA polymerase can be used to convert double-stranded templates into differentiated RNAs that can be used for comparative analysis.

For example, in certain embodiments, a DNA population can be converted to labeled differentiated RNAs by *in vitro* transcription, as shown in FIG. 7.

In the embodiment shown in FIG. 7, multiple nucleic acid populations are differentiated using tag sequences with promoters for different RNA polymerases. In this case, a first nucleic acid tag comprising a differentiation domain comprising a first transcription domain (i.e., a T7 promoter) is appended to a first nucleic acid target of a first nucleic acid population, and a second nucleic acid tag comprising a differentiation domain comprising a second transcription domain (i.e., a SP6 promoter) is appended to the first nucleic acid target of a second nucleic acid population. The differentiation domain of the second nucleic acid tag is different than the differentiation domain of the first nucleic acid tag. In FIG. 7, thick lines represent the tag

sequences, and thin lines represent the sequences of the RNA and/or DNA populations in which one or more nucleic acid targets are comprised.

The differentially tagged samples are mixed to create a sample mixture, and the tagged target(s) in the sample mixture anneal to one or more target specific primers. The concentration of the target specific primer(s) are limited, producing at least a first normalized nucleic acid comprising at least a differentiation domain of the first nucleic acid tag and a nucleic acid segment of the target(s), and at least a second normalized nucleic acid comprising at least a differentiation domain of the second nucleic acid tag and a nucleic acid segment of the target(s).

After the normalization reaction, the sample mixture is split into two differentiation reactions. For example, one differentiation reaction would use T7 RNA polymerase and produce transcripts from the first population, while the second would be transcribed with the SP6 polymerase and produce transcripts from the second population. The resulting transcripts could be analyzed using any of the methods described in this specification.

RNA polymerases are well known to those of ordinary skill in the art. For example, several phage RNA polymerases have been isolated and characterized (Sambrook 1996). Additional RNA polymerases may be isolated from nature or by a mutation/selection screen using an existing polymerase (Ikeda 1993). Any such polymerases or promoters are contemplated for use in the present invention.

2. Differentiation by Restriction Enzyme Digestion

The differentiation domain can also be a site that can be digested to remove a label or affinity group from labeled nucleic acids that constitute a normalized population For example, the differentiation domain can be a restriction site that is digested by a specific restriction enzyme. The digestion can be used to remove an isotopic or non-isotopic label from the termini of normalized nucleic acids from all but one of the sample populations. Alternatively, digestion can displace an affinity group such as biotin from the ends of nucleic acids from one or all but one of the sample populations. The amplified population can then be applied to an avidin support to distinguish members of the amplified population based on whether they had the digestion site within the differentiation domain.

FIG. 8, shows a schematic of such a procedure, whereby multiple nucleic acids can be differentiated using sequences or modifications that that are readily digested. In FIG. 8, a first nucleic acid tag with a differentiation domain comprising a first restriction enzyme domain (i.e., a RE #1) and a labeling agent is appended to a first nucleic acid target of a first nucleic acid population. A second nucleic acid tag with a differentiation domain comprising a second restriction enzyme domain (i.e., a RE #2) and a labeling agent is appended to the first nucleic acid target of a second nucleic acid population The differentiation domain of the second nucleic acid tag is different than the differentiation domain of the first nucleic acid tag. As in other figures in this specification thick lines represent the tag sequences, and thin lines represent the sequences of the RNA and/or DNA populations in which one or more nucleic acid targets are comprised. The differentially tagged samples are mixed to create a sample mixture. The tagged target(s) in the sample mixture anneal to one or more target specific primers, wherein the concentration of the target specific primer(s) are limited, producing at least a first normalized nucleic acid comprising at least a differentiation domain of the first nucleic acid tag and a nucleic acid segment of the target(s), and at least a second normalized nucleic acid comprising at least a differentiation domain of the second nucleic acid tag and a nucleic acid segment of the target(s). The products of primer extension are split into two reactions, one that specifically cleaves products generated from the first sample and one that specifically cleaves products from the second sample. The resulting products are then further analyzed via array or other analysis to reveal the abundance of targets derived from each sample.

3. Affinity sites

A differentiation domain of a tag may comprise a sequence with an affinity for a specific nucleic acid, protein, or other binding ligand. A binding ligand may comprise, but is not limited to, an oligonucleotide complementary to a differentiation domain, a nucleic acid binding protein (*e.g.*, a transcription factor) that bound to a specific DNA or RNA sequence, a small molecule that intercalates into a given RNA or DNA sequence or combinations thereof. A binding ligand may either be bound to a solid support (*e.g.*, a single bead or a membrane in the form of an array) or otherwise readily removed or separated from a solution. An example of this embodiment of the invention is shown in FIG. 9.

As shown in FIG. 9, a first nucleic acid tag with a differentiation domain comprising a first affinity domain (i.e., affinity tag #1) is appended to a first nucleic acid target of a first nucleic acid population. A second nucleic acid tag with a differentiation domain comprising a second affinity domain (i.e., affinity tag #2) is appended to the first nucleic acid target of a second nucleic acid population. The differentiation domain of the second nucleic acid tag is different than the differentiation domain of the first nucleic acid tag. Again, thick lines in FIG. 9 represent the tag sequences, while thin lines represent the sequences of the RNA and/or DNA populations in which one or more nucleic acid targets are comprised. The differentially tagged samples are mixed to create a sample mixture. The tagged target(s) in the sample mixture anneal to one or more target specific primers, wherein the concentration of the target specific primer(s) are limited, producing at least a first normalized nucleic acid comprising at least a differentiation domain of the first nucleic acid tag and a nucleic acid segment of the target(s), and at least a second normalized nucleic acid comprising at least a differentiation domain of the second nucleic acid tag and a nucleic acid segment of the target(s). In specific aspects, a population would then be applied to a solution or solid support possessing a ligand specific to the differentiation domain of one or more of the samples. The specifically isolated targets could then be compared to the unbound or differentially bound nucleic acids.

H. COMPARATIVE ANALYSIS OF LABELED NUCLEIC ACIDS

It is particularly preferred that the combination of limiting competitive normalization and array analysis is used to compare nucleic acid populations. However, other techniques, including ligand-bead separation and chip-based analysis, are preferred in the methods of the present invention to assess normalized target sequences.

1. Arrays

Gene arrays are solid supports upon which a collection of target-specific ligands (*e.g.*, probes) has been attached (*e.g.*, spotted) at defined locations. Arrays for nucleic acid binding are known in the art, and are described, for example, in U.S. Patents 6,156,501, 6,136,962 and 6,087,11, each incorporated herein by reference. The probes may localize complementary nucleic acid (*e.g.*, RNA or DNA) targets from a nucleic acid sample by hybridization. The amount of a target that becomes bound at each spot is a function of the amount of the target present in the sample population. Because the number of ligands that can

be spotted on a gene array is virtually unlimited, arrays can be used to fractionate tens of thousands of target nucleic acid molecules from one or more nucleic acid samples.

Variability in the quality of gene arrays has led many researchers to employ competitive hybridization when comparing different samples (Chee, 1996). Competitive hybridization requires that nucleic acid samples be labeled with distinguishable nucleotides (*e.g.*, two different fluorescent labels). The labeled populations are hybridized to a single array. Labeled targets from the two samples compete for binding to the probes on the array that possess sequences that are complementary to the targets. Following hybridization, the labeled targets bound at each spot on the array are detected. Because the label is unique to sample, the different signals can be quantified separately. Thus, in embodiments wherein different samples are being hybridized to a single array, the amount of target nucleic acid from one sample hybridizing at each target address is a function of the relative abundance of that target in the sample compared to the other sample(s). The method has been extremely effective at reducing experimental variability.

2. Other solid supports

Several methods have been developed that, like array analysis, differentiate targets by sequence-specific hybridization. It is contemplated that these methods, described herein, and others known to those of skill in the art, may be applied in the methods of the present invention, in light of the present disclosures.

a. Beads

Luminex (U.S. Patent 5,981,180), for example, appends target-specific oligonucleotides to fluorescent or otherwise distinguishable beads. Beads for nucleic acid binding are known in the art, and are described, for example, in U.S. Patents 6,133,436, 6,030,782, 5,981,180 and 5,919,626 each incorporated herein by reference. Amplified or differentiated populations (*e.g.*, labeled RNA or DNA populations) can be incubated with the beads to allow target nucleic acids to hybridize to appropriate bead-bound oligonucleotides. The beads are unique for each target, making it possible to simultaneously distinguish target-specific beads and determine the amount of labeled target attached to the bead. The sum of the signal from each target-specific population of beads is totaled, providing a reasonable estimate of the amount of each target in the

sample population. This technology is well-suited for quantifying multiple targets in a population, making it compatible with the present invention.

b. Chip Based Analysis

In another non-limiting example, chip-based formats may also be used to access target nucleic acids by hybridization to oligonucleotides at discrete locations (U.S. Patents 5,632,957, 5,955,028, 5,874,219, 5,861,254 and 5,503,980, each incorporated herein by reference). These methods typically incorporate microchannels etched in silicon wafers. Labeled RNA or DNA populations pass through the channels under the guidance of electrical fields. Interactions between targets and complementary oligonucleotides along these channels localize targets to discrete locations on the chips. The signal from each target specific address provides a reasonable estimate of the concentration of the target in the population. Competitive population normalization could readily be used to generate labeled nucleic acids for analysis in such a format.

I. IDENTIFYING A TAG

Because unique tags are used for different sample populations, it will be very important that the unique tags not contribute to primer extension or differentiation biases (*e.g.*, differences in processing or differentiation efficiencies). The most powerful study contemplated to compare unique tags involves splitting the nucleic acids in a single sample into separate tagging reactions incorporating the different tags. After tagging, the differentially tagged samples can be mixed, normalized, and differentiated. In certain embodiments differentiation comprises generating labeled nucleic acids (*e.g.*, RNA or DNA) using the differentiation domain(s).

The differentiated nucleic acids are assessed by using the method that is to be applied for analysis. For example, if the tags are to be used for array analysis, then the labeled differentiated nucleic acids of the differentiation reaction would be incubated with arrays. If the tags are functioning equivalently, then the probe spots should be identical as they were generated from the same sample population. If signal variation occurs then the tags are biasing the analysis and should be redesigned.

If amplification is to be employed, identifying differentiation domains that function equally well and that do not affect amplification efficiency is relatively straightforward where primer extension or affinity purification is being used for differentiation. In these cases, altering the identity of just a few nucleotides can provide effective differentiation; rarely does altering a few bases within the differentiation domain affect amplification efficiency. In addition, because both methods use the same enzyme (*i.e.*, a single DNA polymerase) for generating labeled nucleic acids polymerization biases should not introduce variability.

If *in vitro* transcription is to be used for labeling (*e.g.*, generating labeled RNA), amplification and/or differentiation bias will be far more likely to occur. Promoters for the well-characterized phage RNA polymerases are similar in base content, but they stretch over 15-20 nucleotides creating a relatively large, unique sequence domain within the amplified nucleic acids. This has the potential to create variability in the amplification reaction. In addition, different RNA polymerases are used for each different tag. Because the different polymerases are likely to possess sequence biases that affect transcription efficiency, the differentiated nucleic acids might not reflect the input samples. This has not affected the method of the present invention in the examples conducted and described herein. However, it is possible that this may affect certain embodiments. To overcome these potential problems, mutants of a single RNA polymerase that do not affect enzymatic activity but do alter promoter specificity may be used in the methods of the present invention may be designed (Ikeda 1993). This methodology may allow the creation of promoter sequences and mutant polymerases that provide equal amplification and differentiation efficiencies to be used to distinguish differentially tagged amplified nucleic acids.

J. EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1

PREPARATION OF ARRAYS TO DISTINGUISH LABELED TARGETS

A collection of RNA standards (Armored RNA constructs IL-1b, IL-2, IL-3, IL-4, IL-5, and IL-7 from Ambion) was heated to 75°C for ten minutes. The RNAs were cooled to room temperature and reverse transcribed in 10 µl reactions containing 5 µM random sequence decamers, reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT), and 20 U/µl MMLV-RT. The cDNA samples were placed in separate PCR™ reactions containing 500 µM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. 200 nM of the target specific primer pairs shown in Table 2 was added to the appropriate reactions and 20 cycles of PCR™ amplified the target inserts.

Table 2
PCR™ primers used to amplify probes for array analysis

IL-1 beta Primers		
SEQ ID NO:1	5' primer:	GGTGTTCTCCATGTCCTTTGT
SEQ ID NO:2	3' primer:	TTGGGGAACTGGGCAGACTCA
IL-2 Primers		
SEQ ID NO:3	5' primer:	CAAACCTCTGGAGGAAGTGCTAA
SEQ ID NO:4	3' primer:	GTGGGAAGCACTTAATTATCAAG
IL-3 Primers		
SEQ ID NO:5	5' primer:	CCTTGTGCGGTTGTGTTCTCATT
SEQ ID NO:6	3' primer:	TCTCACACATCCCTAGGAACCAG
IL-4 Primers		
SEQ ID NO:7	5' primer:	TGCTGCCTCCAAGAACACAACTG
SEQ ID NO:8	3' primer:	CATGATCGTCTTTAGCCTTTCCA
Il-5 Primers		
SEQ ID NO:9	5' primer:	TCGAACTCTGCTGATAGCCAA
SEQ ID NO:10	3' primer:	GCAGTAAAATGTCCTTCTCCTCC
IL-7 Primers		
SEQ ID NO:11	5' primer:	GTGAAGCCCAACCAACAAAGAG
SEQ ID NO:12	3' primer:	TTGGAGGATGCAGCTAAAGTTC

The PCR™ products were denatured in 0.2 M NaOH/10 mM EDTA by heating to 95°C for ten minutes. The denatured PCR™ products were spotted at defined locations on positively charged nylon membranes. The membranes were dried and cross-linked using a StrataLinker set to 120 mJoules.

EXAMPLE 2
TAG SEQUENCES APPENDED TO RNA POPULATIONS BY REVERSE TRANSCRIPTION

This example describes one embodiment wherein a tag comprising a differentiation domain that includes a promoter (*e.g.*, either a T7 or SP6 promoter) was appended to cDNA using reverse transcription.

1 μg of thymus total RNA or liver total RNA was mixed with 0.5 mM dNTPs, 50 μM T7-VN-1 or SP6-VN-1, and reaction buffer (50 mM Tris pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT). The T7-VN-1 and SP6-VN-1 reverse transcription primers were are as follows:

T7-VN-1

GCTGATGGCGATGAATGAACACTG*TAATACGACTCACTATAGGGAG*ATTTTTTTTTTT TVN (SEQ ID NO:13).

SP6-VN-1

GCTGATGGCGATGAATGAACACTG*ATTTAGGTGACACTATAGAAGAG*TTTTTTTTTTT TVN (SEQ ID NO:14).

In regard to these reverse transcription primer sequences:
Bold = Amplification Domain,
Italics = Differentiation Domain, and
Plain Text = Anchored oligodT.

The mixture was heated to 68°C for five minutes, then cooled to 42°C. RNase inhibitor (40 units/μl) and Moloney-Murine Leukemia Virus-Reverse Transcriptase (M-MLV-RT) (20 units/μl) were added and the mixture was incubated at 42°C for one hour to convert the RNA samples to tagged cDNAs.

To remove the reverse transcription primers, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 x G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analyses.

EXAMPLE 3

TAG SEQUENCES APPENDED TO RNA POPULATIONS BY CAP-DEPENDENT LIGATION

This example describes how an RNA can be tagged via cap-dependent ligation. Of course, those of skill in the art will understand that there are additional manners of appending tag sequences, as disclosed in this specification and/or known in the art.

1 μg of mouse thymus or liver total RNA was treated with 1 unit of Calf Intestine Phosphatase in a 20 μl reaction of 50 mM Tris pH 8.5 and 0.1 mM EDTA to remove 5'terminal phosphates from uncapped RNAs. The reaction was incubated at 37°C for 1 hr. The volume of the reaction was increased to 150 μl by the addition of 500 mM Ammonium acetate, 1 mM EDTA. One acid phenol/chloroform and one chloroform extraction were done. The RNA was then precipitated and air dried.

To convert the capped mRNA in the sample to RNA with 5'monophosphates, the dephosphorylated RNA population was dissolved in 8 μl of water. 1 μl of 10X TAP buffer (500 mM NaOAc, 10 mM EDTA, 1% beta-mercaptoethanol, 0.1% CHAPS) and 0.1 unit of TAP were added and the decapping reaction was allowed to proceed for 1 hour at 37°C. Ammonium acetate was added to the reaction to provide a final concentration of 0.5 M. A two-fold volume of ethanol was added and the RNA was precipitated by incubation at -20°C for fifteen minutes. The precipitated RNA was recovered by centrifugation.

To ligate the primers to the RNA possessing a 5' monophosphate, the RNA pellet was dissolved in 8 μl of water. 1 μl of 10X ligation buffer (500 mM Tris pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 1 mM ATP), 1 μg of the ligation tag, and 5 units of T4 RNA ligase were added. The ligation tag sequence is shown, with the amplification domain in bold and the differentiation domain in italics:

5'TAATACGACTCACTATAGGGTTCGGGCTTAGGCTCCAGTGCCTGTTCGGTGGTCGC GGC**GCTGATGGCGATGAATGAACACTG**CGGCAAGCCGCTTAAT*GACACTCGTTTGC TGGCTTTGATG*GGCGAGCTGGAAGGCCGTATCTCCGGCAGCATTCATTACGACAAA 3' (SEQ ID NO:15). The ligation reaction was allowed to proceed for 1 hour at 37°C.

To remove the unincorporated ligation tags, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 x G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

EXAMPLE 4

DISTINGUISHING MIXED POPULATIONS BY PRIMER EXTENSION

Amplified nucleic acids in a sample mixture can be distinguished by extending primers unique to the tag sequences of the different sample populations.

Amplification comprised mixing 1 µl of the cDNA filtrates from Example 2. This mixture was used to initiate an amplification reaction using the primers shown in Table 3.

Table 3

Tag primer (SEQ ID NO:16):	GCTGATGGCGATGAATGAACACTG
Target Specific Primer 1 (SEQ ID NO:17):	GGTGTTCTCCATGTCCTTTGT
Target Specific Primer 2 (SEQ ID NO:18):	CAAACCTCTGGAGGAAGTGCTAA
Target Specific Primer 3 (SEQ ID NO:19):	CCTTGTGCGGTTGTGTTCTCATT
Target Specific Primer 4 (SEQ ID NO:20):	TGCTGCCTCCAAGAACACAACTG
Target Specific Primer 5 (SEQ ID NO:21):	TCGAACTCTGCTGATAGCCAA
Target Specific Primer 6 (SEQ ID NO:22):	GTGAAGCCCAACCAACAAAGAG

The PCR™ comprised 1 µM of the Tag primer and 10 nM of each of the target-specific primers, 500 µM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. PCR™ amplification was by thirty cycles the following profile:

94 °C, 30 seconds; 55 °C, 30 seconds; 72 °C, 60 seconds. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

To remove the amplification primers, the samples were applied to S-200 HR sephacryl spin columns. The columns were spun at 700 x G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

The amplified nucleic acids were diluted into two linear amplification reactions comprising 100 μM dGTP, 100 μM dTTP, 100 μM dCTP, 10 μM dATP, 0.5 μM [α-^{32}P]-dATP, reaction buffer (75 mM KCl, 50 mM TRIS pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. 0.2 μM of Differentiation Primer 1 TAATACGACTCACTATAGGG (SEQ ID NO:23) used in one reaction to synthesize labeled DNA from any reaction products generated from sample 1. A similar amount of Differentiation Primer 2 ATTTAGGTGACACTATAGAA (SEQ ID NO:24) was used to generate differentiated nucleic acids from amplification products derived from sample 2.

Five cycles of the following PCR™ program were used to extend the primer multiple times on the amplified nucleic acid: 94 °C, 5 minutes; then 5 cycles of: 94 °C, 30 seconds; 55 °C, 30 seconds; 72 °C 60 seconds.

EXAMPLE 5
DISTINGUISHING MIXED POPULATIONS BY TRANSCRIPTION

Where the differentiation domains of tags include a transcription promoter, amplified populations can be used to generate labeled RNAs for comparative expression analysis. The procedure described in Example 2 was used to generate cDNA for amplification and transcription to supply populations for comparative expression analysis.

For amplification, 1 μl of the cDNA filtrates from Example 2 were mixed and used to initiate amplification using the primers shown at Table 3. The PCR™ comprised 1 μM of the Tag primer and 10 nM of the target specific primers, 500 μM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. Amplification was

by thirty cycles of PCR™ using the following profile: 94 °C, 30 seconds; 55 °C, 30 seconds; 72 °C, 60 seconds. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

Transcription reactions were then used to convert the amplified nucleic acids into differentiated RNAs for analysis. 2 µl aliquots of the PCR™ samples were used to generate labeled RNAs. Transcription reactions comprising 10 U/µl of T7 or SP6 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM Tris pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 µM UTP, and 2.5 µM [α-^{32}P]-UTP. The reactions were incubated at 37°C for one hour. The reaction with T7 RNA polymerase generated labeled RNA exclusively from amplification products derived from sample 1. The reaction with SP6 RNA polymerase generated labeled RNA exclusively from amplification products derived from sample 2.

EXAMPLE 6

NORMALIZATION OF TARGET SEQUENCES IN A COMPLEX RNA POPULATION

Two sample nucleic acid populations were constructed using four distinct RNAs. The RNAs included sequence from IL-3, IL-4, IL-5 and IL-7. The samples contained different amounts of each RNA as indicated below.

Target	Copies in Population #1	Copies in Population #2
IL-3	10^{10}	10^{11}
IL-4	10^{9}	10^{9}
IL-5	10^{8}	10^{7}
IL-7	10^{7}	10^{6}

The two RNA samples were reverse transcribed using the following tag primer using the protocol described in Example 2:

Amplification Domain *Differentiation Domain*

GCTGATGGCGATGAATGAACACTG*TAATACGACTCACTATAGGGAGA* TTTTTTTTTTTTVN (SEQ ID NO:25)

Unincorporated primers were removed by applying the samples to S-200 HR sephacryl spin columns. The columns were spun at 700 × G for 2 minutes. The filtrate was recovered, providing the tagged cDNA population used for subsequent analysis.

Amplification was conducted by mixing 1 µl of the cDNA filtrates and initiating amplification using the tag primer and the target specific primers 3, 4, 5 and 6, shown at Table 3. The PCR™ comprised 1 µM of the Tag primer and 0.1 nM of the target specific primers, 500 µM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. Amplification was by thirty cycles of PCR™ using the following profile: 94 °C, 30 seconds; 55 °C, 30 seconds; 72 °C, 60 seconds. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

To generate labeled RNA, 2 µl aliquots of the two PCR™ samples were used in a transcription reaction. Transcription reactions comprising 10 U/µl of T7 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM Tris pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 µM UTP, and 2.5 µM [α-^{32}P]-UTP. The reactions were incubated at 37°C for one hour.

The labeled RNAs were prepared for array analysis by heat denaturing in 100 µl of 10 mM EDTA. The RNAs were added to the arrays described in Example 1 that had been prehybridized for 30 minutes at 42°C in hybridization buffer (2.5X SSC, 50% formamide, 7% SDS, 200 µg/ml yeast RNA). The blots were hybridized overnight at 42°C in hybridization buffer, then washed 2 X 30" at 42°C in 2X SSC/0.5% SDS and 2 X 30" at 60°C in 0.5X SSC/0.5% SDS. The washed blots were exposed overnight to film. The resulting autoradiographs provided approximately equal signal from each of the spots on the array from both sample populations. In contrast, arrays hybridized with samples that had been amplified with 100 nM of each of the target specific primers failed to detect IL-5 and IL-7 in the second RNA sample.

EXAMPLE 7
COMPETITIVE POPULATION NORMALIZATION

Where the differentiation domains of tags include a transcription promoter, amplified populations can be used to generate labeled RNAs for comparative expression analysis. The procedure described in Example 2 was used to generate cDNA for amplification and transcription to supply populations for comparative expression analysis.

For amplification, 1 µl of the cDNA filtrates of example 1 was used with the tag primer and target specific primers 3, 4, 5 and 6, shown in Table 3. The PCR™ comprised 1 µM of the Tag primer and 100 pM of the target specific primers, 500 µM dNTPs, reaction buffer (75 mM KCl, 50 mM Tris pH 9.0, 2 mM $MgCl_2$), and 2.5 units SuperTaq polymerase. The targets were amplified by PCR™ using 30 cycles of the following PCR™ profile: 94 °C, 30 seconds; 55 °C, 30 seconds; 72 °C, 60 seconds. At the end of the final amplification cycle, a 5 minute, 72°C soak was used for annealing and primer extension.

Transcription reactions using 2 µl aliquots of the PCR™ samples were used to generate labeled RNAs. Transcription reactions comprising 10 U/µl of T7 or SP6 RNA polymerase, reaction buffer (20 mM NaCl, 40 mM Tris pH 8, 6 mM $MgCl_2$, 2 mM spermidine HCl, and 10 mM DTT), 0.5 mM ATP, GTP, and CTP, 10 µM UTP, and 2.5 µM [α-^{32}P]-UTP. The reactions

were incubated at 37°C for one hour. The use of the two polymerases depended on the identity of the tag used for reverse transcription. The cDNAs primed by T7-VN-1 were transcribed with T7 RNA polymerase. The cDNAs primed by SP6-VN-1 were transcribed with SP6.

The RNAs were added to the arrays described in Example 1 that had been prehybridized for 30 minutes at 42°C in hybridization buffer (2.5X SSC, 50% formamide, 7% SDS, 200 μg/ml yeast RNA). The blots were hybridized overnight at 42°C in hybridization buffer, then washed 2 × 30" at 42°C in 2X SSC/0.5% SDS and 2 X 30" at 60°C in 0.5X SSC/0.5% SDS. The washed blots were exposed overnight to film. The resulting autoradiographs provided approximately equal signal from each of the spots on the array from both sample populations. This was as expected as the cDNAs derived from the same RNA sample and had been normalized during the amplification reaction.

* * *

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

Beasley *et al.*, "Statistical refinement of primer design parameters," *In: PCR Applications*, Innis, Gelfand, Sninsky (Eds.), pp. 55-72, 1999.

Butler and Chamberlin, *J. Biol. Chem.*, 257:5772-5778, 1982.

Chamberlin and Ryan, in The Enzymes, ed. P. Boyer Academic Press, New York, pp. 87-108, 1982.

Chapman and Burgess, *Nucleic Acids Res.* 16:5413, 1987.

Chapman and Wells, *Nucleic Acids Res.* 10(20):6331, 1982.

Chee *et al*, "Assessing genetic information with high density DNA Arrays," Science, 274:610-614, 1996.

Compton J. "Nucleic acid sequence-based amplification," *Nature.* 7;350(6313):91-92, 1991.

Diaz *et al.*, *J. Mol. Biol.* 229: 805-811, 1993.

Duggan *et al.*, "Expression profiling using cDNA microarrays," Nat. Genet., 21:S10-S14, 1999.

Dunn and Studier, *J. Mol. Biol.* 166:477-535, 1983; and erratum *J. Mol. Biol.* 175:111-112, 1984.

Dunn *et al.*, *Nature New Biology*, 230:94-96, 1971.

Egholm *et al.*, "PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules," *Nature*, 365(6446):566-568, 1993.

EPO App. No. 266,032

EPO App. No. 329 822

EPO App. No. 98302726

Froehler *et al.*, "Synthesis of DNA via deoxynucleoside H-phosphonate intermediates," *Nucleic Acids Res.* 14(13):5399-5407, 1986.

GB Application 2 202 328

Guatelli *et al.*, "Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication," *Proc Natl Acad Sci U S A.* 87(5):1874-1878, 1990.

Hausmann, *Current Topics in Microbiology and Immunology*, 75:77-109, 1976.

Henegariu *et al.*, "Multiplex PCR: critical parameters and step-by-step protocol," *Biotechniques.* 23(3):504-511, 1997.

Ikeda *et al.*, "Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences," Biochem., 32:9115-9124, 1993.

Innis *et al.*, "DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA," *Proc. Natl. Acad. Sci. USA*, 85(24):9436-9440, 1988.

Kato, "Adaptor-tagged competitive-PCR™: A novel method for measuring relative gene expression," Nuc. Acids Res., 25:4694-4696, 1997.

Klement, *et al.*, *J. Mol. Biol.* 215:21-29, 1990.

Kornberg, A. and T. A. Baker, DNA Replication, Second Edition, 1992, New York, W. H. Freeman and Company.

Korsten *et al.*, *J. Gen. Virol.*, 43:57-73, 1975.

Kotani *et al.* (1987), *Nucl. Acids Res.* 15:2653-2664, 1987.

Kwoh *et al.*, "Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format, *Proc. Natl. Acad. Sci. USA*, 86(4):1173-1177, 1989.

Lipshutz *et al.*, "High density synthetic oligonucleotide arrays," Nat. Genet., 21:S20-S24, 1999.

Lizardi *et al.*, *Bio/technology* 6:1197-1202, 1988.

Lockhart *et al.*, "Expression monitoring by hybridization to high-density oligonucleotide arrays," *Nat Biotechnol.* 14(13):1675-1680, 1996.

Lomeli *et al.*, *Clin. Chem.* 35:1826-1831, 1989

Matoba *et al.*, *Gene* 241:125-131,2000

Morris *et al.*, *Gene* 41:221-227, 1986

PCT Application EP/01219

PCT Application US89/01025

PCT Application WO 88/10315

Phillips and Eberwine, Methods: A Companion to Methods in Enzymology 10:283-288, 1996.

Sambrook *et al.*, *In: Molecular Cloning: A laboratory Manual,* Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Ch. 7,7.19-17:29, 1989.

Scheit, Nucleotide Analogs, Synthesis and Biological Function, Wiley-Interscience, 1980, New York, pp. 171-172.

Schneider and Stormo, G. D., *Nucleic Acids Res.* 17(2):659, 1989.

Towle *et al.*, *J. Biol. Chem.*, 250:1723-1733, 1975.

U.S. Patent 4,659,774

U.S. Patent 4,682,195

U.S. Patent 4,683,195

U.S. Patent 4,683,202

U.S. Patent 4,786,600

U.S. Patent 4,800,159
U.S. Patent 4,816,571
U.S. Patent 4,952,496
U.S. Patent 4,959,463
U.S. Patent 5,141,813
U.S. Patent 5,214,136
U.S. Patent 5,223,618
U.S. Patent 5,264,566
U.S. Patent 5,279,721
U.S. Patent 5,340,728
U.S. Patent 5,378,825
U.S. Patent 5,428,148
U.S. Patent 5,446,137
U.S. Patent 5,466,786
U.S. Patent 5,470,967
U.S. Patent 5,514,545
U.S. Patent 5,539,082
U.S. Patent 5,545,522
U.S. Patent 5,554,744
U.S. Patent 5,574,146
U.S. Patent 5,602,240
U.S. Patent 5,602,244
U.S. Patent 5,610,289
U.S. Patent 5,614,617
U.S. Patent 5,623,070
U.S. Patent 5,632,957
U.S. Patent 5,645,897
U.S. Patent 5,652,099
U.S. Patent 5,670,663
U.S. Patent 5,672,697
U.S. Patent 5,681,947
U.S. Patent 5,700,922

U.S. Patent 5,705,629
U.S. Patent 5,708,154
U.S. Patent 5,714,331
U.S. Patent 5,714,606
U.S. Patent 5,719,262
U.S. Patent 5,736,336
U.S. Patent 5,763,167
U.S. Patent 5,766,855
U.S. Patent 5,773,571
U.S. Patent 5,777,092
U.S. Patent 5,786,461
U.S. Patent 5,792,847
U.S. Patent 5,824,528
U.S. Patent 5,830,694
U.S. Patent 5,840,873
U.S. Patent 5,843,640
U.S. Patent 5,843,650
U.S. Patent 5,843,651
U.S. Patent 5,843,708
U.S. Patent 5,843,717
U.S. Patent 5,846,709
U.S. Patent 5,846,726
U.S. Patent 5,846,729
U.S. Patent 5,846,783
U.S. Patent 5,849,487
U.S. Patent 5,849,497
U.S. Patent 5,849,546
U.S. Patent 5,849,547
U.S. Patent 5,853,990
U.S. Patent 5,853,992
U.S. Patent 5,853,993
U.S. Patent 5,856,092

U.S. Patent 5,858,652
U.S. Patent 5,859,221
U.S. Patent 5,861,244
U.S. Patent 5,863,732
U.S. Patent 5,863,753
U.S. Patent 5,866,331
U.S. Patent 5,866,366
U.S. Patent 5,872,232
U.S. Patent 5,882,864
U.S. Patent 5,886,165
U.S. Patent 5,891,625
U.S. Patent 5,891,681
U.S. Patent 5,905,024
U.S. Patent 5,908,845
U.S. Patent 5,910,407
U.S. Patent 5,912,124
U.S. Patent 5,912,145
U.S. Patent 5,916,776
U.S. Patent 5,919,630
U.S. Patent 5,922,574
U.S. Patent 5,925,517
U.S. Patent 5,928,006
U.S. Patent 5,928,862
U.S. Patent 5,928,869
U.S. Patent 5,928,905
U.S. Patent 5,929,227
U.S. Patent 5,932,413
U.S. Patent 5,932,451
U.S. Patent 5,935,791
U.S. Patent 5,935,825
U.S. Patent 5,939,291
U.S. Patent 5,942,391

U.S. Patent 5,955,028

U.S. Patent 5,962,271

U.S. Patent 5,981,180

U.S. Patent 6,015,664

U.S. Patent 6,025,134

U.S. Patent 6,057,134

U.S. Patent 6,107,037

Watson *et al.*, Molecular biology of the Gene, 4th Ed., Chapters 13-15, Benjamin/Cummings Publishing Co., Menlo Park, Calif. (1988).

WO 90/07641

WO 92/20702

CLAIMS

1. A method for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each potentially having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of at least a first target specific primer to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein any first product nucleic acids comprise the first differentiation domain and a section of the first nucleic acid target from

the first sample and any second product nucleic acids comprise the second differentiation domain and a section of the first nucleic acid target from the second sample;

g) differentiating any first product nucleic acids from any second product nucleic acids; and

h) comparing the amount of first nucleic acid target in the first sample, if any, to the amount of first nucleic acid target of the second sample, if any.

2. The method of claim 1, wherein the processing of the sample mixture further comprises amplification to produce amplified nucleic acid from the first sample and amplified nucleic acid from the second sample.

3. The method of claim 2, wherein the amplification process comprises a plurality of primer extension reactions.

4. The method of claim 2, wherein the first tag and the second tag both comprise amplification domains.

5. The method of claim 4, wherein a primer specific to the amplification domain is added to the extension reaction.

6. The method of claim 5, wherein a non-limiting amount of a primer specific to the amplification domains and a limited amount of a primer specific to the first target are added to the sample mixture and the processing of the sample mixture comprises multiple cycles of primer extension to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample.

7. The method of claim 4, wherein the amplification domain of the first nucleic acid tag and the second nucleic acid tag comprises a primer binding domain.

8. The method of claim 4, wherein the amplification domain of the first nucleic acid tag and the second nucleic acid tag comprises a transcription domain.

9. The method of claim 4, wherein the amplification domains of the first and second nucleic acid tags are functionally equivalent.

10. The method of claim 9, wherein the amplification domains of the first and second nucleic acid tags are identical.

11. The method of claim 4, wherein the differentiation domain of the first tag and the second tag is appended between the at least a first nucleic acid target sequence and the amplification domain.

12. The method of claim 1, wherein the first nucleic acid target is one target of a plurality of nucleic acid targets within the samples.

13. The method of claim 1, wherein the first and second sample are two samples of a plurality of samples.

14. The method of claim 13, wherein the first and second tag are two tags of a plurality of tags.

15. The method of claim 1, wherein the first target specific primer is one of a plurality of target specific primers, each target specific primer present in a limiting concentration.

16. The method of claim 15, wherein the first target specific primers anneal to the first target nucleic acid.

17. The method of claim 1, wherein the differentiation domain of the first nucleic acid tag or the second nucleic acid tag comprises a primer binding domain, a transcription domain, an affinity domain, a restriction enzyme cleavage domain, or a combination thereof.

18. The method of claim 1, wherein processing comprises production of at least one first product nucleic acid complementary to the first nucleic acid target of the first sample.

19. The method of claim 1, wherein the differentiation domain of the first nucleic acid tag and the differentiation domain of the second nucleic acid tag each comprise a distinct primer binding domain.

20. The method of claim 19, wherein the differentiating comprises binding at least a first differentiation primer to at least one segment of the primer binding domain.

21. The method of claim 20, further comprising at least one primer extension reaction.

22. The method of claim 21, wherein the primer extension reaction produces at least one differentiated nucleic acid.

23. The method of claim 22, wherein the differentiated nucleic acid is labeled with a detectable moiety.

24. The method of claim 1, wherein the differentiation domain of the first nucleic acid tag and the differentiation domain of the second nucleic acid tag each comprise distinct transcription domains.

25. The method of claim 24, wherein the differentiation domain comprises an prokaryotic RNA polymerase promoter.

26. The method of claim 24, wherein the differentiating comprises at least one transcription reaction.

27. The method of claim 26, wherein the transcription reaction produces at least one differentiated nucleic acid.

28. The method of claim 27, wherein the differentiated nucleic acid is labeled with a detectable moiety.

29. The method of claim 1, wherein the differentiation domain of the first nucleic acid tag or the second nucleic acid tag comprises at least one affinity domain.

30. The method of claim 29, wherein the differentiating comprises binding at least a first ligand to at least a segment of the affinity domain.

31. The method of claim 30, wherein the first ligand comprises a nucleic acid.

32. The method of claim 30, wherein the first ligand is labeled with a detectable moeity.

33. The method of claim 30, wherein the first ligand is bound to at least one solid support.

34. The method of claim 33, wherein the solid support comprises an array, a microtiter well, a chip, a glass slide, a bead or a combination thereof.

35. The method of claim 33, wherein the bound nucleic acids are removed from the first ligand.

36. The method of claim 1, wherein the differentiation domain of the first nucleic acid tag or the second nucleic acid tag comprises at least one restriction enzyme cleavage domain.

37. The method of claim 36, further comprising cleaving the restriction enzyme cleavage site to promote the ligation of a label to a segment of the at least a first or at least a second nucleic acid tag sequence.

38. The method of claim 37, further comprising cleaving the restriction enzyme cleavage site to promote the ligation of at least one additional domain to a segment of the at least a first or at least a second nucleic acid tag sequence.

39. The method of claim 36, wherein the differentiating comprises cleaving the restriction enzyme site to remove at least one label.

40. The method of claim 1, wherein the first nucleic acid tag or the second nucleic acid tag further comprises at least one additional domain.

41. The method of claim 40, wherein the additional domain is a labeling domain, a restriction enzyme domain, a secondary amplification domain, a secondary differentiation domain or a combination thereof.

42. A method for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of first target specific primers to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction with the mixed sample to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, wherein the first product nucleic acids comprise the first differentiation domain and a section of the first nucleic acid target from the first sample and the second product nucleic acids comprise the second differentiation domain and a section of the first nucleic acid target from the second sample;

g) differentiating the first product nucleic acids from the second product nucleic acids; and

h) comparing the amount of first nucleic acid target in the first sample, to the amount of first nucleic acid target of the second sample.

43. A method for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each potentially having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, wherein the first differentiation domain comprises a first transcription domain;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second

sample, wherein the second differentiation domain comprises a second transcription domain that is different than the first transcription domain;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of first target specific primers to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein any first product nucleic acids comprise the first transcription domain and a section of the first nucleic acid target from the first sample and any second product nucleic acids comprise the second transcription domain and a section of the first nucleic acid target from the second sample;

g) differentiating any first product nucleic acids complementary to the first nucleic acid target of the first sample from any second product nucleic acids complementary to the first nucleic acid target of the second sample by a process comprising at least a first transcription reaction from the first transcription domain to produce an amount of at least first differentiated nucleic acid, if the first nucleic acid target is present in the first sample, and a second transcription reaction from the second transcription domain to produce at least second differentiated nucleic acid, if the first nucleic acid target is present in the second sample; and

h) comparing the amount of the first nucleic acid target in the first sample, if any, with the amount of the first nucleic acid target in the second sample, if any, by comparing the first differentiated nucleic acid, if any, to the amount of the second differentiated nucleic acid, if any.

44. A method for comparing one or more nucleic acid targets within two or more samples comprising:

a) obtaining at least a first sample and a second sample, each potentially having at least a first nucleic acid target;

b) preparing at least a first tagged nucleic acid sample by appending at least a first nucleic acid tag comprising a first differentiation domain to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, wherein the first differentiation domain comprises a first primer binding domain;

c) preparing at least a second tagged nucleic acid sample by appending at least a second nucleic acid tag comprising a second differentiation domain to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein the second differentiation domain comprises a second primer binding domain that is different than the first primer binding domain;

d) mixing the first tagged nucleic acid sample and the second tagged nucleic acid sample to create a sample mixture;

e) adding a limiting concentration of first target specific primers to the sample mixture;

f) processing the sample mixture by a process comprising at least a first extension reaction to produce a limited concentration of first product nucleic acids complementary to the first nucleic acid target of the first sample, if the first nucleic acid target is present in the first sample, and a limited concentration of second product nucleic acids complementary to the first nucleic acid target of the second sample, if the first nucleic acid target is present in the second sample, wherein any first product nucleic acids comprise the first primer binding domain and a section of the first nucleic acid target

from the first sample and any second product nucleic acids comprise the second primer binding domain and a section of the first nucleic acid target from the second sample;

g) differentiating any first product nucleic acids complementary to the first nucleic acid target of the first sample from any second product nucleic acids complementary to the first nucleic acid target of the second sample by a process comprising annealing at least a first differentiation primer to the first primer binding domain and performing at least a second primer extension reaction to produce an amount of at least first differentiated nucleic acid, if the first nucleic acid target is present in the first sample, and annealing at least a second differentiation primer to the second primer binding domain and performing at least a third primer extension reaction to produce an amount of at least second differentiated nucleic acid, if the first nucleic acid target is present in the second sample; and

h) comparing the amount of the first nucleic acid target in the first sample, if any, with the amount of the first nucleic acid target in the second sample, if any, by comparing the first differentiated nucleic acid, if any, to the amount of the second differentiated nucleic acid, if any.

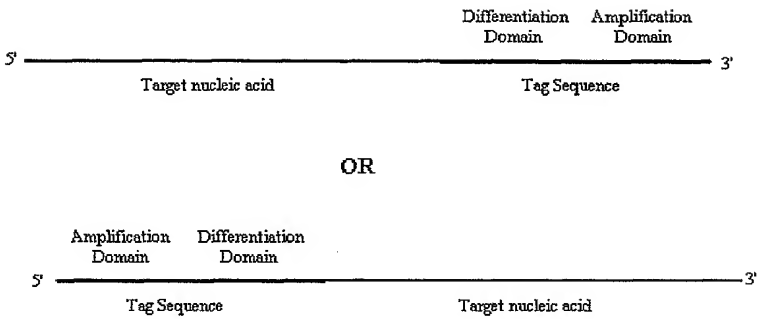
Differentiation Domain
Amplification Domain
5'
3'
Target nucleic acid
Tag Sequence
OR
Amplification Domain
Differentiation Domain
5'
3'
Tag Sequence
Target nucleic acid

FIG. 1

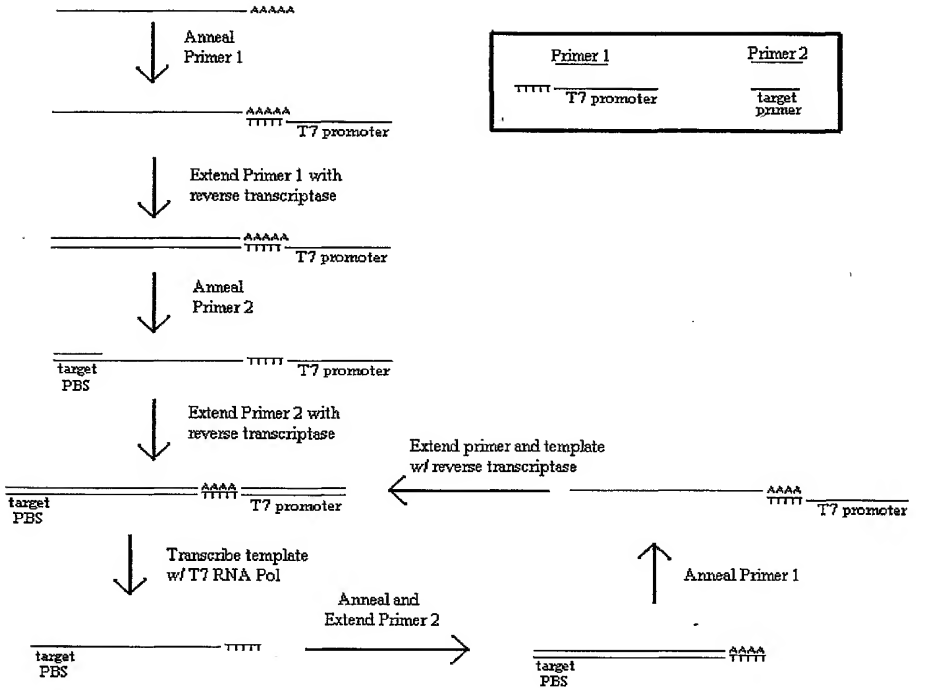
Anneal
Primer 1
T7 promoter
Extend Primer 1 with
reverse transcriptase
Anneal
Primer 2
target
PBS
Extend Primer 2 with
reverse transcriptase
Extend primer and template
w/ reverse transcriptase
Transcribe template
w/ T7 RNA Pol
Anneal and
Extend Primer 2
Anneal Primer 1
Primer 1
Primer 2
target
primer

FIG. 2

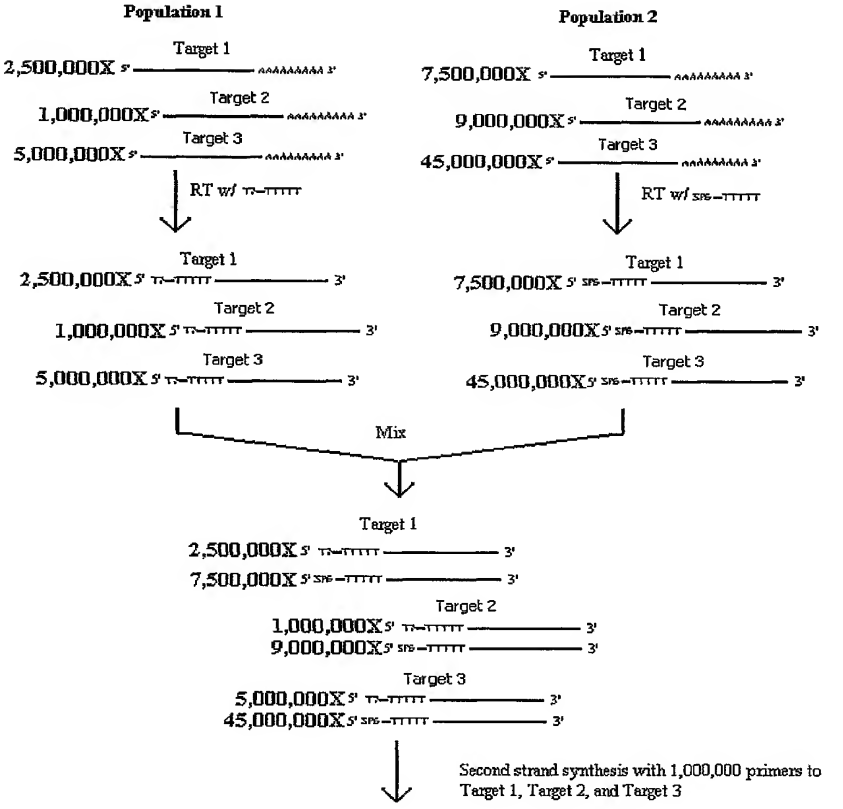
Population 1
Target 1
2,500,000X
Target 2
1,000,000X
Target 3
5,000,000X
RT w/
Target 1
2,500,000X
Target 2
1,000,000X
Target 3
5,000,000X
Population 2
Target 1
7,500,000X
Target 2
9,000,000X
Target 3
45,000,000X
RT w/
Target 1
7,500,000X
Target 2
9,000,000X
Target 3
45,000,000X
Mix
Target 1
2,500,000X
7,500,000X
Target 2
1,000,000X
9,000,000X
Target 3
5,000,000X
45,000,000X
Second strand synthesis with 1,000,000 primers to
Target 1, Target 2, and Target 3

FIG. 3A

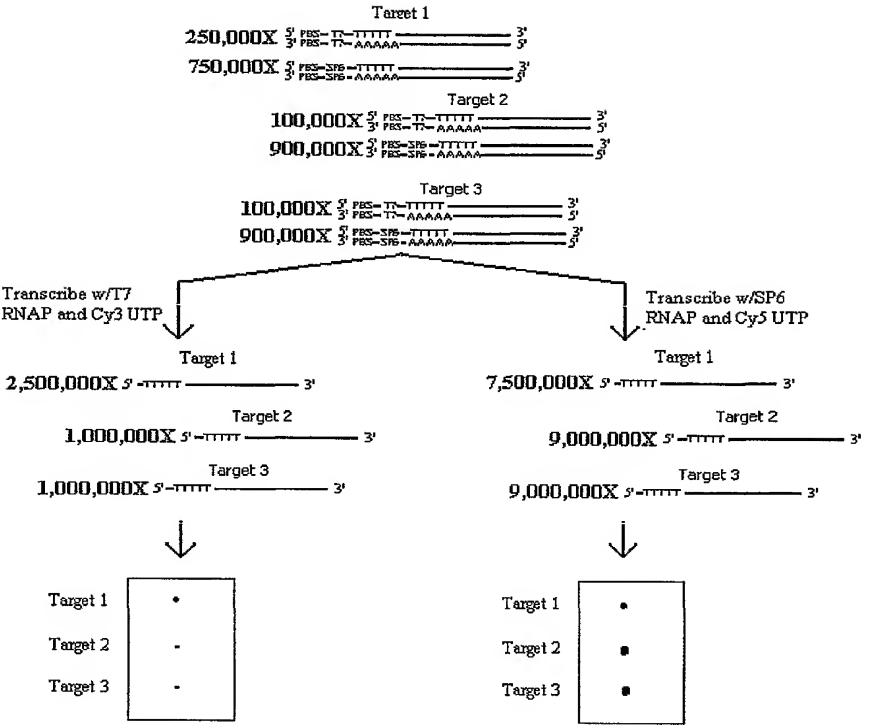
Target 1
250,000X
750,000X
Target 2
100,000X
900,000X
Target 3
100,000X
900,000X
Transcribe w/T7 RNAP and Cy3 UTP
Transcribe w/SP6 RNAP and Cy5 UTP
Target 1
2,500,000X
Target 2
1,000,000X
Target 3
1,000,000X
Target 1
7,500,000X
Target 2
9,000,000X
Target 3
9,000,000X
Target 1
Target 2
Target 3
Target 1
Target 2
Target 3

FIG. 3B

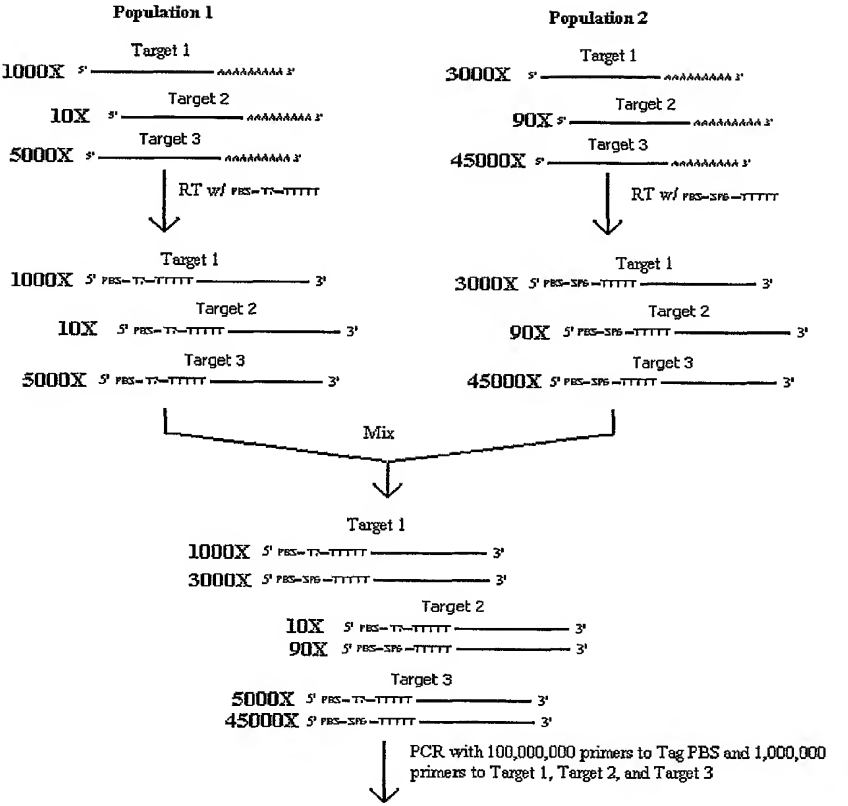
Population 1
Target 1
1000X
Target 2
10X
Target 3
5000X
RT w/ PBS-T7-TTTTT
Target 1
1000X 5' PBS-T7-TTTTT 3'
Target 2
10X 5' PBS-T7-TTTTT 3'
Target 3
5000X 5' PBS-T7-TTTTT 3'
Population 2
Target 1
3000X
Target 2
90X
Target 3
45000X
RT w/ PBS-SP6-TTTTT
Target 1
3000X 5' PBS-SP6-TTTTT 3'
Target 2
90X 5' PBS-SP6-TTTTT 3'
Target 3
45000X 5' PBS-SP6-TTTTT 3'
Mix
Target 1
1000X 5' PBS-T7-TTTTT 3'
3000X 5' PBS-SP6-TTTTT 3'
Target 2
10X 5' PBS-T7-TTTTT 3'
90X 5' PBS-SP6-TTTTT 3'
Target 3
5000X 5' PBS-T7-TTTTT 3'
45000X 5' PBS-SP6-TTTTT 3'
PCR with 100,000,000 primers to Tag PBS and 1,000,000 primers to Target 1, Target 2, and Target 3

FIG. 4A

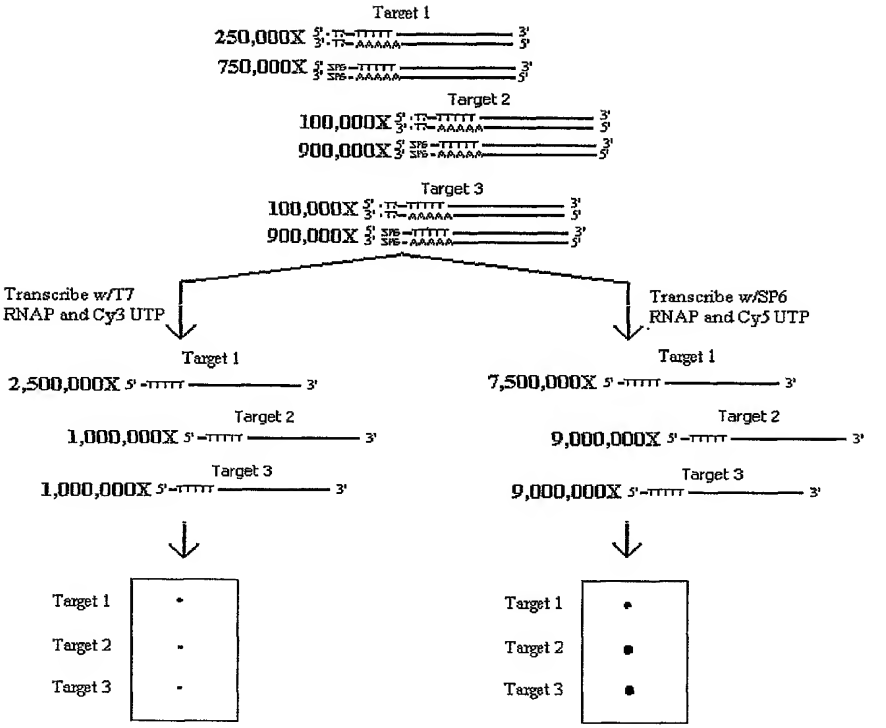
Target 1
250,000X
750,000X
Target 2
100,000X
900,000X
Target 3
100,000X
900,000X
Transcribe w/T7 RNAP and Cy3 UTP
Transcribe w/SP6 RNAP and Cy5 UTP
Target 1
2,500,000X
Target 2
1,000,000X
Target 3
1,000,000X
Target 1
7,500,000X
Target 2
9,000,000X
Target 3
9,000,000X
Target 1
Target 2
Target 3
Target 1
Target 2
Target 3

FIG. 4B

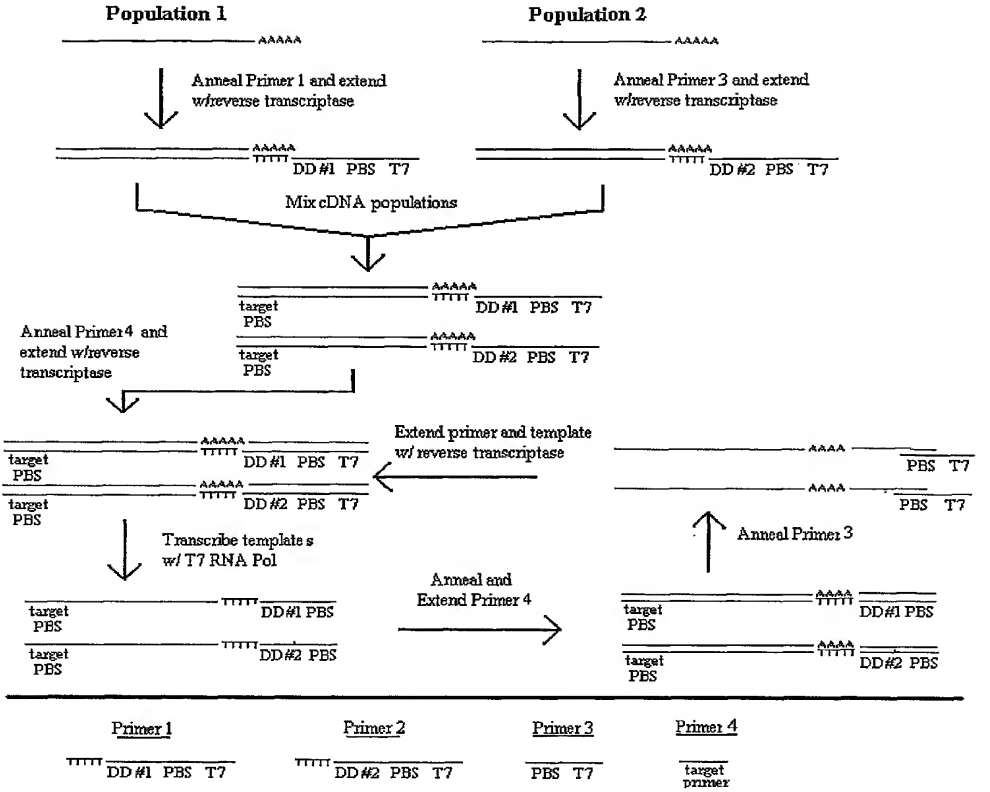
Population 1
Population 2
Anneal Primer 1 and extend w/reverse transcriptase
Anneal Primer 3 and extend w/reverse transcriptase
DD#1 PBS T7
DD#2 PBS T7
Mix cDNA populations
target PBS
Anneal Primer 4 and extend w/reverse transcriptase
Extend primer and template w/ reverse transcriptase
Transcribe template s w/ T7 RNA Pol
Anneal and Extend Primer 4
Anneal Primer 3
DD#1 PBS
DD#2 PBS
Primer 1
Primer 2
Primer 3
Primer 4
PBS T7
target primer

FIG. 5

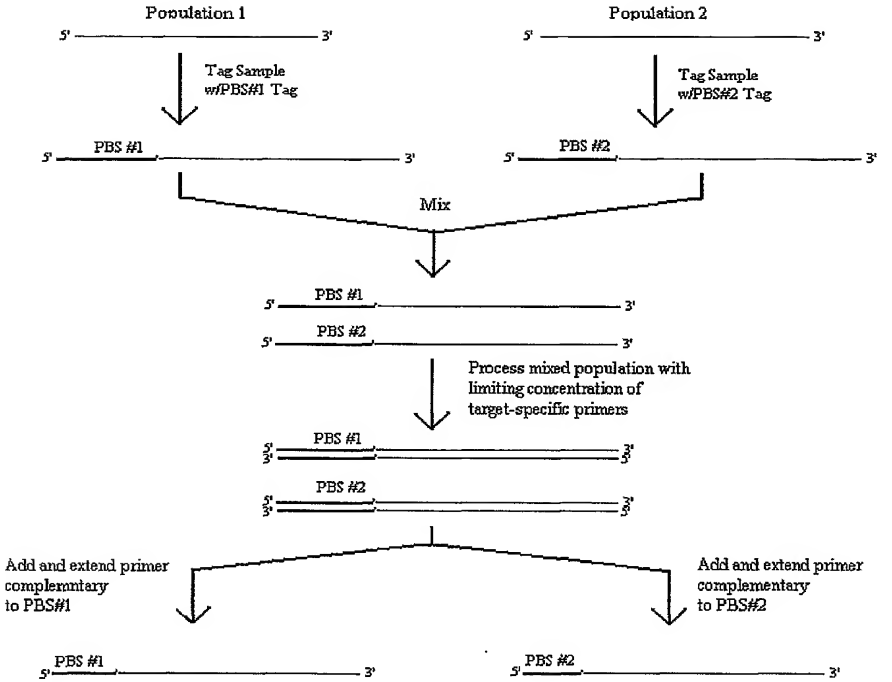
Population 1
5'
3'
Tag Sample
w/PBS#1 Tag
PBS #1
Population 2
Tag Sample
w/PBS#2 Tag
PBS #2
Mix
Process mixed population with
limiting concentration of
target-specific primers
Add and extend primer
complementary
to PBS#1
Add and extend primer
complementary
to PBS#2

FIG. 6

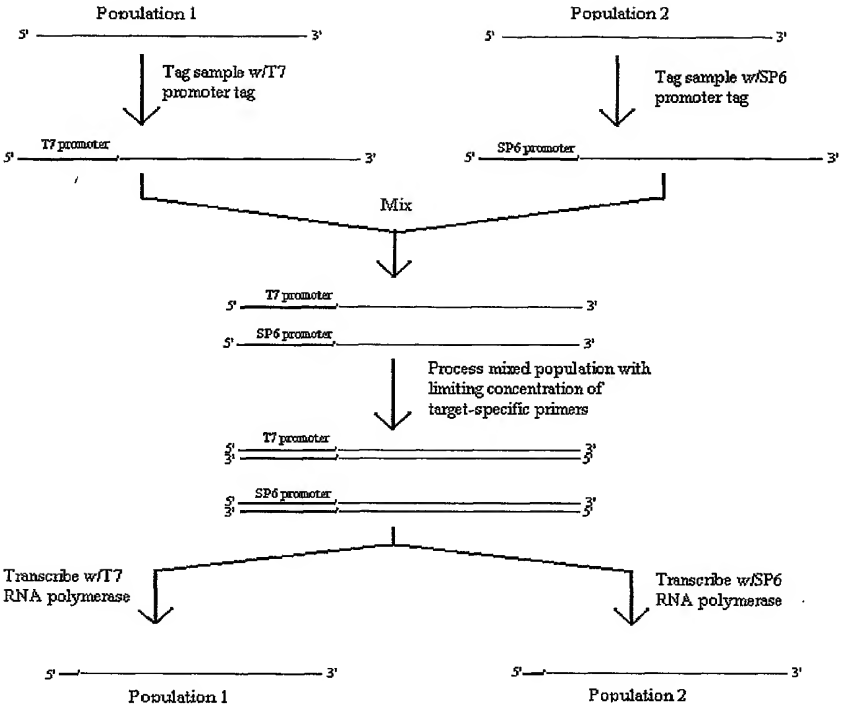
Population 1
Population 2
Tag sample w/T7 promoter tag
Tag sample w/SP6 promoter tag
T7 promoter
SP6 promoter
Mix
Process mixed population with limiting concentration of target-specific primers
Transcribe w/T7 RNA polymerase
Transcribe w/SP6 RNA polymerase
Population 1
Population 2

FIG. 7

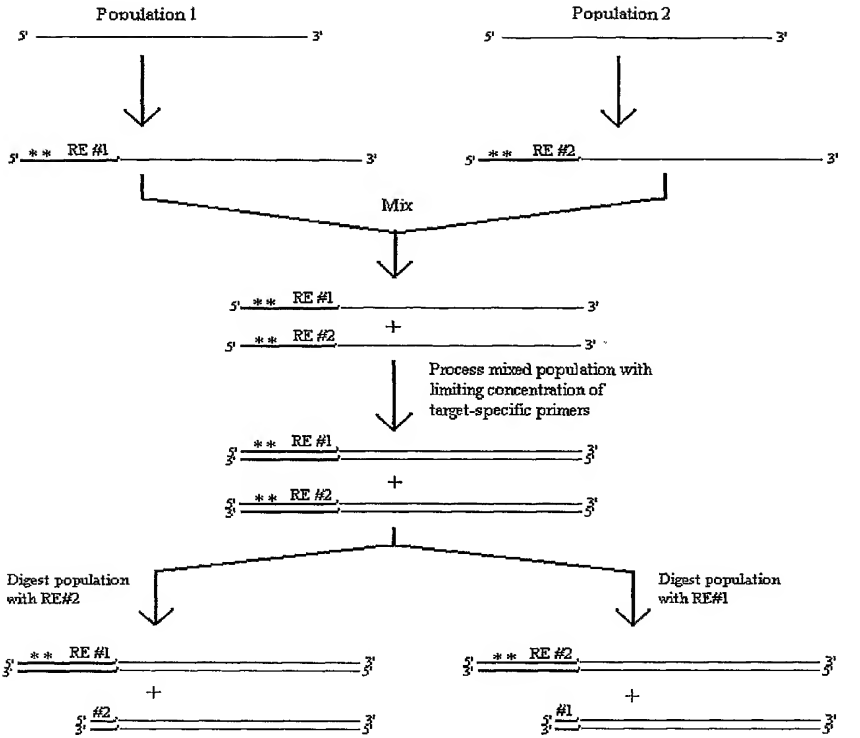
Population 1
Population 2
RE #1
RE #2
Mix
Process mixed population with limiting concentration of target-specific primers
Digest population with RE#2
Digest population with RE#1
#2
#1

FIG. 8

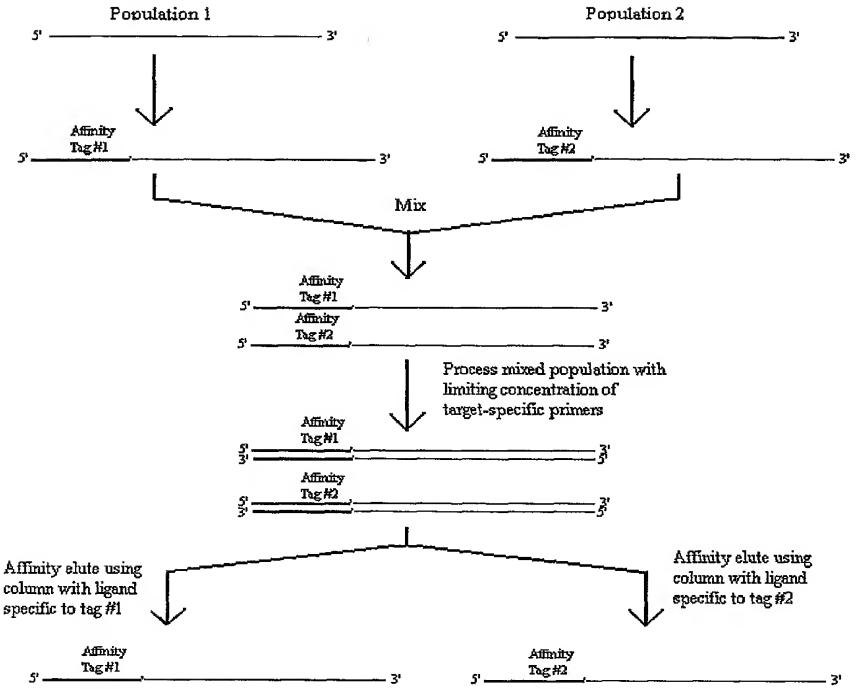
Population 1
Population 2
Affinity Tag #1
Affinity Tag #2
Mix
Process mixed population with limiting concentration of target-specific primers
Affinity elute using column with ligand specific to tag #1
Affinity elute using column with ligand specific to tag #2

FIG. 9

SEQUENCE LISTING

<110> BROWN, DAVID
WINKLER, MATTHEW M.

<120> COMPETITIVE POPULATION NORMALIZATION FOR COMPARATIVE ANALYSIS OF NUCLEIC ACID SAMPLES

<130> AMBI:064-WO

<140> UNKNOWN
<141> 2002-01-31

<150> 60/265,265
<151> 2001-01-31

<160> 25

<170> PatentIn Ver. 2.1

<210> 1
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 1
ggtgttctcc atgtcctttg t 21

<210> 2
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 2
ttggggaact gggcagactc a 21

<210> 3
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 3
caaacctctg gaggaagtgc taa 23

<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 4
gtgggaagca cttaattatc aag 23

<210> 5
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 5
ccttgtgcgg ttgtgttctc att 23

<210> 6
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 6
tctcacacat ccctaggaac cag 23

<210> 7
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 7
tgctgcctcc aagaacacaa ctg 23

<210> 8
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 8
catgatcgtc tttagccttt cca 23

<210> 9
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 9
tcgaactctg ctgatagcca a 21

<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 10
gcagtaaaat gtccttctcc tcc 23

<210> 11
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 11
gtgaagccca accaacaaag ag 22

<210> 12
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 12
ttggaggatg cagctaaagt tc 22

<210> 13
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<220>
<221> modified_base
<222> (60)
<223> V =G, C or A

<400> 13
gctgatggcg atgaatgaac actgatttag gtgacactat agaagagttt ttttttttv 60
n 61

<210> 14
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<220>
<221> modified_base
<222> (60)
<223> V = G, C or A

<400> 14
gctgatggcg atgaatgaac actgatttag gtgacactat agaagagttt ttttttttv 60
n 61

<210> 15
<211> 168
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 15
taatacgact cactataggg ttcgggctta ggctccagtg cctgttcggt ggtcgcggcg 60
ctgatggcga tgaatgaaca ctgcggcaag ccgcttaatg acactcgttt gctggctttg 120
atgggcgagc tggaaggccg tatctccggc agcattcatt acgacaaa 168

<210> 16
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 16
gctgatggcg atgaatgaac actg 24

<210> 17
<211> 24
<212> DNA
<213> Artificial Sequence

<400> 17
gctgatggcg atgaatgaac actg 24

<210> 18
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 18
caaacctctg gaggaagtgc taa 23

<210> 19
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 19
ccttgtgcgg ttgtgttctc att 23

<210> 20
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 20
tgctgcctcc aagaacacaa ctg 23

<210> 21
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 21
tcgaactctg ctgatagcca a 21

<210> 22
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 22
gtgaagccca accaacaaag ag 22

<210> 23
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 23
taatacgact cactataggg 20

<210> 24
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 24
atttaggtga cactatagaa 20

<210> 25
<211> 61
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<220>
<221> modified_base
<222> (61)
<223> V = G, C or A

<400> 25
gctgatggcg atgaatgaac actgtaatac gactcactat agggagattt ttttttttv 60
n 61